

ADA025319

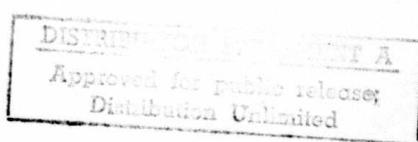
*Preliminary Reports, Memoranda
and Technical Notes of the
Materials Research Council
Summer Conference*

La Jolla, California

July 1975

**Sponsored by
Advanced Research Projects Agency
ARPA Order No. 2341/2**

Department of Materials and Metallurgical Engineering



PRELIMINARY REPORTS, MEMORANDA AND TECHNICAL NOTES
of the
MATERIALS RESEARCH COUNCIL SUMMER CONFERENCE
La Jolla, California

Jul 1975

(12) 687 p. /

ARPA Order Number: 2341/2
Program Code Number: 1D10
Contractor: The Regents of The University of Michigan
Effective Date of Contract: 1 May 75
Contract Expiration Date: 30 June 76
Amount of Contract: \$265,000
Contract Number: DAHC15-71-C-0253, ~~ARPA Order~~
Principal Investigator: Associate Dean Maurice J. Sinn

15 10 College of Engineering
The University of Michigan
Ann Arbor Michigan 48109

10 College of Engineering
The University of Michigan
Ann Arbor, Michigan 48109
(313) 763-0242

ACCESSION for		
NTIS	White Section <input checked="" type="checkbox"/>	
DOC	Buff Section <input type="checkbox"/>	
UNNUMBERED	<input type="checkbox"/>	
JUSTIFICATION	<input type="checkbox"/>	
BY		
DISSEMINATION AND SECURITY CODES		
DISSEM	SECURITY	TYPE OF SPECIAL
A		

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or the U.S. Government.

TABLE OF CONTENTS

- I. Foreword
- II. Steering Committee
- III. Participants
- IV. Consultants
- V. Preliminary Reports, Memoranda and Technical Notes

The following papers fall into two categories; (1) papers in a state ready for publication, and (2) reports and memoranda for limited distribution representing work in progress. The former category is available for general distribution and in some cases are in the process of publication in the appropriate technical journals. The limited distribution reports and memoranda represent initial ideas, problem suggestions, position papers, and status reports and are aimed primarily to stimulate discussion with the Council. However, they are available subject to the author's release by request to the Project Director.

<u>TITLE</u>	<u>PAGE</u>
One- and Two-Dimensional Conductors W. H. Flygare, G. D. Stucky and H. Ehrenreich . . .	1
Antenna Structures for Infrared Diodes P. L. Richards and A. Yariv	65
Degradation of Photovoltaic Cells W. Kohn	87
Theoretical Issues Related to Solids Containing Highly Mobile Atoms or Ions R. A. Huggins	95
Higher Order Processes in Tunneling Barrier Diodes A. Yariv.	129
Summary of Meeting on Optical Electronics N. Bloembergan.	143
Fundamental Considerations Limiting Tunnel Diode Mixers and Detectors M. Tinkham.	155

<u>TITLE</u>	<u>PAGE</u>
Optical Band Gaps of Oxides and Sulphides J. J. Gilman.	185
Wear Workshop A. Bienenstock, B. Budiansky, D. C. Drucker J. P. Hirth, H. Reiss and J. R. Rice.	193
Quasi-Static Solid Particle Damage in Brittle Solids A. G. Evans	205
Contribution on the Issues of Wear Particle Size, Composition and Structure A. W. Ruff.	247
Comparison between the Adhesive and Fatigue Explanations of Adhesive Wear E. Rabinowicz	251
Ion Implantation and Wear J. W. Butler.	255
Comments on the Bowden and Tabor Asperity Theory of Wear V. Westcott	271
Fundamentals of Wear A. H. Shabaik	277
On the Delamination Theory of Wear N. P. Suh	283
Fatigue Analysis of Wear B. Budiansky and F. A. McClintock	291
Elementary Results of Dimensional Analysis for Rates of Wear in Steady State Sliding D. C. Drucker	299
The Role of Crystal Plasticity in the Delamination Theory of Wear J. P. Hirth	309
Wear, Polishing and the Study of Surface Layers by X-ray Diffraction A. Bienenstock.	327
Surface Mechanics and Wear F. F. Ling.	339

<u>TITLE</u>	<u>PAGE</u>
Erosion by Solid Particles I. Finnie	361
The Thermodynamic Criterion for Crack Growth R. M. Thomson	383
Notes on the Mechanics and Thermodynamics of Brittle Interfacial Failure in the Presence of a Mobile Species (Hydrogen) J. R. Rice.	397
Materials/Design A. G. Evans	413
Research on Material Characterization for Structural Reliability and Design J. R. Rice.	423
The Materials/Design Interface R. M. Thomson and M. Cohen.	437
Analysis of the Duwez Splat-Cooling Process M. Tinkham.	457
Summary of the Meeting on Forming Shapes of Advanced Materials from Powders E. E. Hucke	469
Toward a Coordinated Program on Powder-Processed Materials E. E. Hucke, J. P. Hirth, M. Cohen and P. Duwez . .	475
On the Powder Forming of Advanced Materials M. Cohen.	481
Quantitative Nondestructive Evaluation M. Cohen and R. M. Thomson.	491
Non-Destructive Evaluation of Ceramics G. S. Kino and A. G. Evans.	513
Coal Liquification as an Intercalation Problem R. A. Huggins	535
Synthesis of Large Diamonds for Optical Components J. J. Gilman.	545
On the Design of Alloys for Strength and Fracture Toughness M. Cohen.	553

<u>TITLE</u>	<u>PAGE</u>
Concentration Fluctuations in Adsorbed Layers Brian Bell, R. Gomer and H. Reiss	565
Feasibility of Operating of Energy Conversion Machines Under Reducing Conditions at Very High Temperatures E. E. Hucke	597
Acid Rainwater J. L. Margrave.	599
An Approach for Preparing Antifouling Concrete for Undersea Use C. U. Pittman, Jr.	607
A Qualitative Approach to Anisotropic Semiconducting Polymeric Materials C. U. Pittman, Jr.	617
Chemical Anchoring of Mildewcides in Paints: An Analysis and Discussion of Its Feasibility C. U. Pittman, Jr.	629
Comments on Synthetic Approaches to One-Dimensional Conductors C. U. Pittman, Jr.	665
Molecular Rectifiers? C. U. Pittman, Jr.	669

Foreword

This collection of papers does not constitute a formal reporting of the activities of the ARPA Materials Research Council Summer Conference. Each report, memorandum or technical note is a draft of the author or authors and is their work alone. The Steering Committee, in conjunction with the authors, will decide how this material can best be presented as a formal report to ARPA.

Steering Committee

Professor Willis H. Flygare
Secretary of the Steering Committee
Department of Chemistry
University of Illinois
Urbana, Illinois 61801

Professor Nico Bloembergen
Division of Engineering & Applied Physics
Harvard University
Cambridge, Massachusetts 02138

Professor Bernard Budiansky
Division of Engineering & Applied Science
Harvard University
Cambridge, Massachusetts 02138

Professor Morris Cohen
Department of Metallurgy & Materials Science
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Dr. John J. Gilman, Director
Materials Research Center
Allied Chemical Corporation
Morristown, New Jersey 07960

Professor Paul L. Richards
Department of Physics
University of California
Berkeley, California 94720

Dr. Robb M. Thomson
National Bureau of Standards
Institute for Materials Research
Washington, D.C. 20234

Project Director

Associate Dean Maurice J. Sinnott
College of Engineering
The University of Michigan
Ann Arbor, Michigan 48109

Participants

Professor Arthur Bienenstock
Materials Science Department
Stanford University
Stanford, California 94305

Professor Robert Coble
Materials Science Department
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Dean Daniel C. Drucker
Engineering College
University of Illinois
Urbana, Illinois 61801

Dr. Anthony G. Evans
Rockwell International Science Center
Thousand Oaks, California 91360

Professor Henry Ehrenreich
Pierce Hall
Harvard University
Cambridge, Massachusetts 02138

Professor Robert Gomer
James Franck Institute
University of Chicago
Chicago, Illinois 60637

Professor M. Fred Hawthorne
Department of Chemistry
University of California
Los Angeles, California 90024

Professor John P. Hirth
Metallurgical Engineering Department
Ohio State University
Columbus, Ohio 43201

Professor Robert A. Huggins
Center for Materials Research
Stanford University
Stanford, California 94305

Professor Gordon S. Kino
Department of Electrical Engineering
Stanford University
Stanford, California 94305

Professor Walter Kohn
Department of Physics
University of California
La Jolla, California 92037

Professor John L. Margrave
Department of Chemistry
Rice University
Houston, Texas 77001

Professor Frank A. McClintock
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Professor Elliott W. Montroll
Department of Physics & Astronomy
University of Rochester
Rochester, New York 14534

Professor Howard Reiss
Department of Chemistry
University of California
Los Angeles, California 90024

Professor James Rice
Division of Engineering
Brown University
Providence, Rhode Island 02912

Professor Michael Tinkham
Department of Physics
Harvard University
Cambridge, Massachusetts 02138

Dr. George H. Vineyard
Brookhaven National Laboratory
Upton, Long Island, New York 11973

Professor Amnon Yariv
Electrical Engineering Department
California Institute of Technology
Pasadena, California 91109

GUEST CONSULTANTS

A. M. Adair
AFML/LLM
Wright Patterson AFB
Ohio 45433

R. E. Allen
General Electric Company
Evendale, Ohio 45231

P. W. Anderson
Bell Telephone Laboratories
Murray Hill, New Jersey 07974

B. Avitzur
Lehigh University
Bethlehem, Pennsylvania 13015

R. H. Baughman
Allied Chemical Corporation
Morristown, New Jersey 07960

F. D. Bedard
Lab. for Physical Sciences
College Park, Maryland 20740

M. J. Blackburn
Pratt & Whitney Aircraft
East Hartford, Conn. 06037

E. D. Blaustein
U.S. Bureau of Mines
Pittsburgh, Pennsylvania 15213

A. N. Bloch
Dept. of Chemistry
Johns Hopkins University
Baltimore, Maryland 21218

R. G. Brandt
Office of Naval Research
Pasadena, California

R. H. Bube
Dept. of Materials Science
Stanford University
Stanford, California 94305

J. W. Butler
Naval Research Laboratory
Code 6670
Washington, D.C. 20375

A. Cohn
Electric Power Research Inst.
Palo Alto, California 94303

A. R. Cox
Pratt & Whitney Aircraft
W. Palm Beach, Florida

R. deWit
National Bureau of Standards
Washington, D.C. 20234

A. M. Diness
Office of Naval Research
Arlington, Virginia 22217

F. J. DiSalvo
Bell Telephone Laboratories
Murray Hill, New Jersey 07974

J. Doi
Chemistry Department
University of California
Los Angeles, California 90024

W. H. Dukes
Bell Aerospace Company
New Orleans, Louisiana 70189

P. E. Duwez
California Institute of Tech.
Pasadena, California 91125

C. Elbaum
Physics Department
Brown University
Providence, Rhode Island 02912

E. Edelsack
Office of Naval Research
Arlington, Virginia 22217

K. Evenson
National Bureau of Standards
Boulder, Colorado 80302

I. Finnie
University of California
Berkeley, California 94705

S. Flatté
Physics Department
University of California
Santa Barbara, California

T. Garito
University of Pennsylvania
Philadelphia, Pennsylvania

I. S. Goldstein
Dept. of Wood & Paper Science
North Carolina State University
Raleigh, North Carolina 27607

N. J. Grant
Director, Materials Science
Center
Massachusetts Inst. of Tech.
Cambridge, Massachusetts 02139

R. Green
IBM
San Jose, California

J. Greenspan
Army Materials & Mechanics
Research Center
Watertown, Massachusetts

H. M. Grotta
 Battelle Memorial Institute
Columbus, Ohio 43201

T. K. Gustafson
University of California
Berkeley, California 94705

B. M. Harney
ERDA
Washington, D.C. 20455

A. J. Heeger
Department of Physics
University of Pennsylvania
Philadelphia, Pennsylvania 19174

H. Heinemann
Mobil Research & Dev. Corp.
Princeton, New Jersey 08540

J. R. Herbert
AAS Corporation
Baltimore, Maryland 21264

A. Hurlich
Convair Div. General Dynamics
El Cajon, California 92020

A. Javan
Mechanical Engineering Dept.
Massachusetts Inst. of Tech.
Cambridge, Massachusetts 02139

J. J. Jimenez
National Bureau of Standards
Boulder, Colorado 80302

R. E. Johnson
NASA-Johnson Space Center
Houston, Texas

W. C. Kalb
Chemistry Department
University of California
Los Angeles, California 90024

K. J. Klabunde
Department of Chemistry
University of North Dakota
Grand Forks, North Dakota 58202

G. Y. Lai
General Atomic Company
San Diego, California

W. J. Levedahl
Naval Ship Research & Dev. Ctr.
Annapolis, Maryland 21402

M. Levy
Army Materials & Mechanics
Research Center
Watertown, Massachusetts 02172

F. F. Ling
Rensselaer Polytechnic Inst.
Troy, New York 12181

H. A. Lipsitt
Air Force Materials Lab.
Wright-Patterson AFB
Ohio

N. H. MacMillan
Pennsylvania State Univ.
University Park, Pennsylvania
16802

A. Maimoni
Lawrence Livermore Laboratory
Livermore, California

H. I. McHenry
National Bureau of Standards
Boulder, Colorado 80302

T. E. Miles
Kelsey Hayes Company
Brighton, Michigan 48116

R. S. Miller
Office of Naval Research
Arlington, 22217

P. B. Mohr
Lawrence Livermore Laboratory
Livermore, California 94550

J. Moore
Pratt & Whitney Aircraft
W. Palm Beach, Florida

B. Myers
Lawrence Livermore Laboratory
Livermore, California 94550

G. W. Parshall
Central Research Department
DuPont Company
Wilmington, Delaware 19898

C. U. Pittman
Chemistry Department
University of Alabama
University, Alabama 35486

E. Rabinowicz
Mechanical Engineering Dept.
Massachusetts Inst. of Tech.
Cambridge, Massachusetts 02139

R. Reed
National Bureau of Standards
Boulder, Colorado 80302

R. Reynolds
Materials Science Office
Advanced Research Projects Agency
Arlington, Virginia 22209

J. F. Roth
Monsanto Company
St. Louis, Missouri 63166

L. Roth
Hughes Research Laboratories
Malibu, California 90265

S. Ruby
Materials Science Office
Advanced Research Projects Agency
Arlington, Virginia 22209

W. Ruff
National Bureau of Standards
Washington, D.C. 20234

A. Sanchez
Massachusetts Inst. of Tech.
Cambridge, Massachusetts 02139

H. Schlossberg
AFCRL
Hanscom AFB, Maryland 01731

A. Schriesheim
Exxon Research & Engineering Co.
Linden, New Jersey

S. E. Schwarz
Electrical Engineering Dept.
University of California
Berkeley, California 94720

B. A. Scott
IBM T. J. Watson Research
Center
Yorktown Heights, New York 10598

R. P. Sernka
Aeronutronic Ford Corp.
Newport Beach, California 92663

D. E. Severson
Dept. of Chemical Engineering
University of North Dakota
Grand Forks, North Dakota 58202

A. Shabaik
Dept. of Mechanical Engineering
University of California
Los Angeles, California 90024

O. D. Sherby
Materials Science & Engineering
Stanford University
Stanford, California 94305

A. H. Silver
Aerospace Corporation
Los Angeles, California 90009

P. H. Stahlhuth
Aeronutronic-Ford
Newport Beach, California 92663

F. W. Steffgan
ERDA
Pittsburgh Energy Research Ctr.
Pittsburgh, Pennsylvania 15213

C. M. Stickley
Materials Science Office
Advanced Research Projects Agency
Arlington, Virginia 22209

G. D. Stucky
Chemistry Department
University of Illinois
Urbana, Illinois 61801

N. P. Suh
Dept. of Mechanical Engineering
Massachusetts Inst. of Tech.
Cambridge, Massachusetts 02139

P. Tannenwald
M.I.T. Lincoln Laboratories
Lexington, Massachusetts 02173

J. R. Tucker
Aerospace Corporation
Los Angeles, California

A. van der Ziel
Electrical Engineering Dept.
University of Minnesota
Minneapolis, Minnesota 55455

E. C. van Reuth
Materials Science Office
Advanced Research Projects Agency
Arlington, Virginia 22209

S. Wang
Dept. of Electrical Engineering
University of California
Berkeley, California 94720

N. K. Welker
Laboratory for Physical Sciences
College Park, Maryland 20740

E. T. Wessel
Westinghouse Electric Corp.
Research & Development Center
Pittsburgh, Pennsylvania 15235

V. C. Westcott
Foxboro/Trans-Sonics Inc.
Burlington, Massachusetts 01803

S. Wiederhorn
National Bureau of Standards
Washington, D.C. 20234

J. J. Wilczynski
Chemistry Department
University of California
Los Angeles, California 90024

E. H. S. Wong
Chemistry Department
University of California
Los Angeles, California 90024

S. Yukawa
General Electric Company
Schenectady, New York 12345

ONE- AND TWO-DIMENSIONAL CONDUCTORS

W. H. Flygare
G. D. Stucky
H. Ehrenreich

- I. INTRODUCTION
- II. ONE-DIMENSIONAL MATERIALS
 - A. Transition Metal Bridged Systems
 - B. Square Planar Transition Metal (d^8) Complexes
 - C. Organic Charge Transfer Salts
 - D. Polymeric Conductors
- III. TWO-DIMENSIONAL AND LAYERED SYSTEMS
- IV. APPLICATIONS
- V. CONCLUSIONS AND RECOMMENDATIONS

APPENDIX

ONE- AND TWO-DIMENSIONAL CONDUCTORS

W. H. Flygare
G. D. Stucky
H. Ehrenreich

I. INTRODUCTION

This report summarizes the material presented and discussed at a two day meeting organized by the ARPA Materials Research Council on One- and Two-Dimensional Conductors which was held in La Jolla, California, on July 10 and 11, 1975 (an agenda for the meeting is shown in the Appendix). For reference the electronic conductivity of a number of systems is shown in the following diagram.

The meeting was organized approximately as follows:

A. General summary of generic inorganic and organic systems which have been examined. General summary of the chemical and physical principle involved in designing one-dimensional molecular conductors.

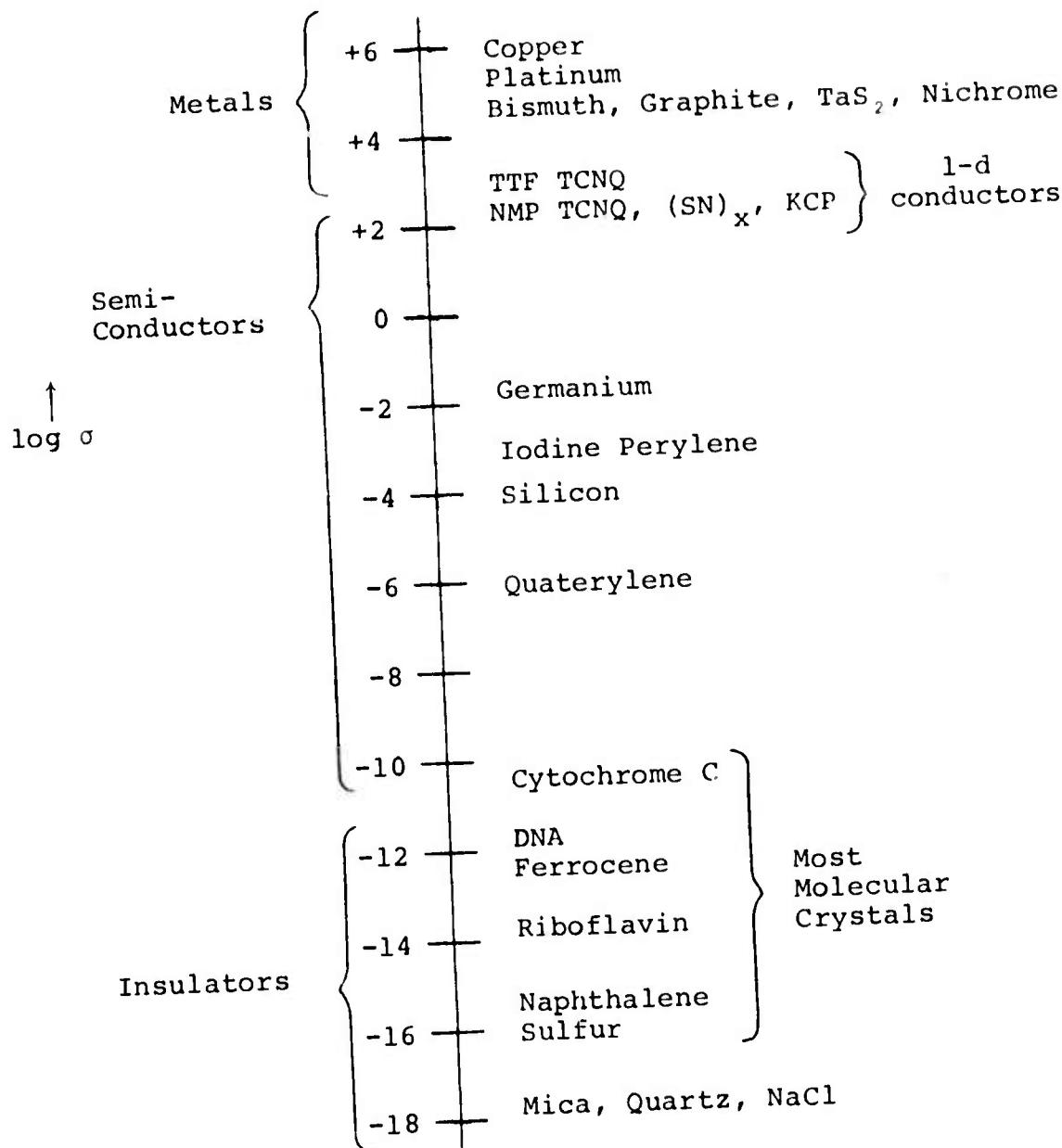
B. Full examination of the extensively studied charge-transfer complexes with particular emphasis on TTF-TCNQ as a model compound which has received a great deal of attention.

C. Polymeric conductors with special interest on the physical features of $(SN)_x$.

D. Studies of two-dimensional and layered compound electronic conductors.

E. Technological applications of highly anisotropic one- and two-dimensional conductors.

CONDUCTIVITY SCALE (25°C)



The report which follows is organized essentially as outlined above.

II. ONE-DIMENSIONAL MATERIALS

The relatively large family of one-dimensional (1-d) materials can be broken down into four major classifications:

(1) ligand-bridged transition metal complexes, (2) square planar transition metal compounds derived from the d^8 electronic configuration of the metal atom, (3) organic charge transfer complexes, and (4) polymeric conductors. By far, the majority of these materials are insulators or semiconductors, and only a few generic systems have been found to exhibit metallic behavior.

In fact, it should be pointed out that the metallic 1-d materials now being extensively investigated are derivatives of compounds which were discovered many years ago. $[K_2Pt(CN)_4]Br_{0.30} \cdot 3H_2O$ was first prepared by Knop in 1842¹ and $(SN)_x$ was isolated and characterized by Burt in 1910². Metallic ($\sigma > 10^2 \text{ ohm}^{-1}\text{cm}^{-1}$) 1-d organic charge transfer TCNQ complexes with stacks of like donor molecules and separate stacks of acceptor molecules were first prepared at DuPont in the early 1960's and characterized by single crystal structural and conductivity studies by 1965^{3,4}.

It appears unlikely that distinctively new classes of compounds not included above will be discovered. Thus, it is important to invent and examine new materials in the above classes with the varied mechanical and electronic properties needed for practical applications. These synthetic efforts are

also essential in order to furnish the background necessary for a systematic and predictive understanding of the physical behavior of 1- and 2-d materials.

There are however many difficulties in synthesizing new systems. In order to have cooperative electronic behavior, one must have relatively large interatomic or intermolecular interactions. For the most part this means materials which are not very volatile, which are insoluble in most solvents, and which frequently decompose before melting. A material with these properties is particularly difficult to purify, characterize and crystallize. In addition, the requirement that there be strong intermolecular and interatomic interactions requires that non-bonding steric interactions between atoms and molecules be minimized. A final added chemical complication is that the ingredients for such materials are free radical species which, with the exceptions of transition metal complexes and a few organic radicals, are notoriously reactive. Normally there are many products which can be formed which have a smaller entropy of formation than an infinite one-dimensionally ordered chain. Assuming that a chemist is sufficiently clever to overcome the above difficulties, the physics of one- and two-dimensional metals suggests that they should be inherently more unstable than three dimensional materials. Instabilities associated with a large intra-electronic correlation energy in narrow band systems leading to Mott transitions^{5,6} and potential perturbations which introduce a band gap at the Fermi level through a soft phonon mode

leading to a Peierls distortion⁷ are important in this regard. The Peierls distortion may be particularly important in limiting the number of one dimensional materials which can be readily synthesized.

Lee, Rice and Anderson⁸ have suggested that the condensation energy for a charge density wave (CDW) associated with a half filled one dimensional band may be comparable to the commensurability energy of the material that it may be impossible to achieve a stable metallic state. Indeed, there are currently no known examples of 1-d metallic materials with half filled bands. In $(SN)_x$, formally a half filled band structure, the existence of interactions in directions perpendicular to the chain structure indicates substantial 3-d character. The charge transfer in TTF-TCNQ is estimated to be 0.45 ± 0.15 of one electron by the direct integration of the electron density measured by single crystal X-ray diffraction studies, 0.55 from X-ray neutron inelastic scattering studies⁹, and 0.56-0.67 from photo emission studies. These results indicate a $\sim 1/4$ filled band in TTF-TCNQ. In the mixed valence d⁸ systems, the conduction band contains ~ 1.70 electrons. Known linear chain systems with half filled bands such as $NbCl_4$ and $[Pd^{II}(NH_3)_2Br_2-Br-Pd^{IV}(NH_3)_2Br_2-]$ can be considered to be examples of 1-d Peierls distorted molecules and are insulators or semiconductors. These observations, in fact put a stringent requirement on the chemist, since it means that a compound which is a free radical with one unpaired electron in a non-degenerate orbital probably will not form a

one-dimensional metal. The restrictions in two dimensions are much less stringent, but nevertheless may be limiting.

From the standpoint of furthering the field of one- and two-dimensional materials, it is important that the synthetic chemist have a thorough understanding of the requirements for designing such systems. This is particularly so since, as pointed out above, the probability of accidentally discovering such systems is small.

In this section a brief structural and electronic description of some examples in each of the four classifications described above will be presented. This will be followed by a more detailed description of the low temperature physics and transport properties of some specific examples.

A. Transition Metal Bridged Ligand Systems

Many attempts have been made by inorganic chemists during the past few years to devise 1-d conducting systems in which transition metal atoms are constrained to a one-dimensional configuration by bridging ligands. Some examples of these geometries are shown in Figures 1 and 2. Probably primarily because of the difficulty in obtaining sufficiently large metal-metal transfer integrals, this approach has been successful with only one known system, BaVS_3 , which is structurally one dimensional (Figure 3) and which has a 300°K bulk conductivity of more than $10^2 \text{ ohm}^{-1}\text{cm}^{-1}$.¹⁰ The conductivity is temperature dependent and BaVS_3 appears to undergo a metal-insulator transition at 75°K with an accompanying change in crystal structure symmetry from

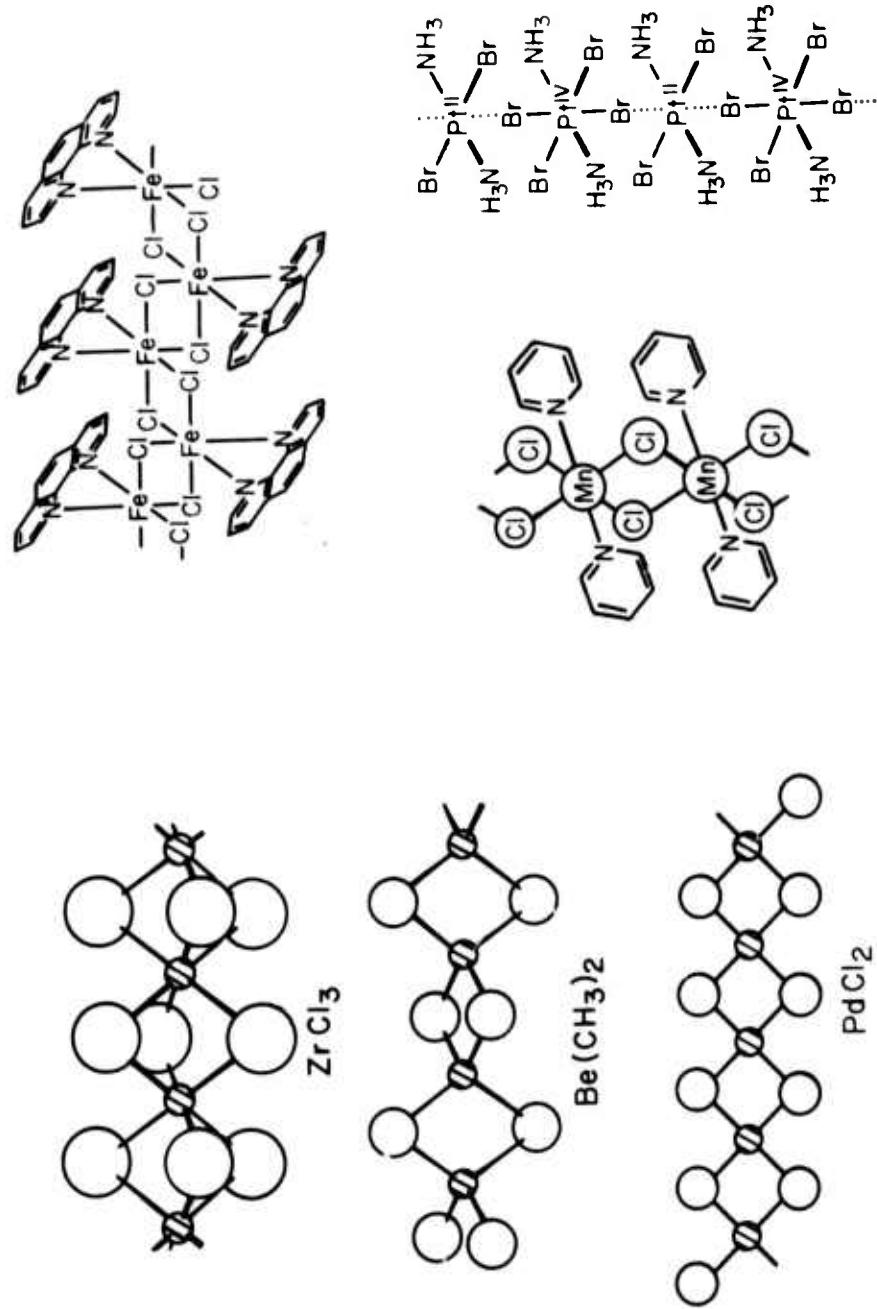


Figure 1. Examples of Structurally 1-d Materials with Single Atom Bridging.

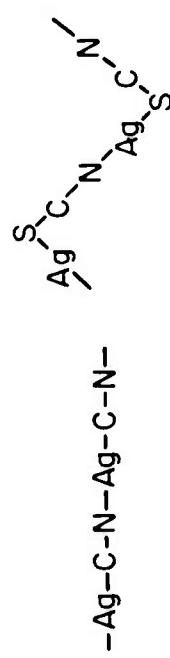
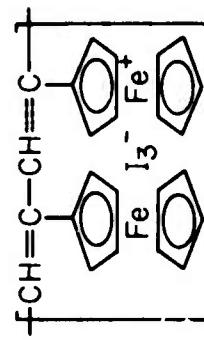
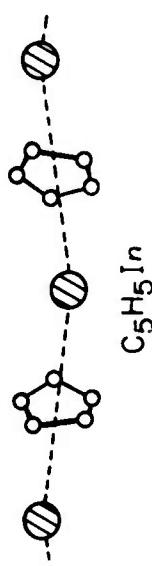
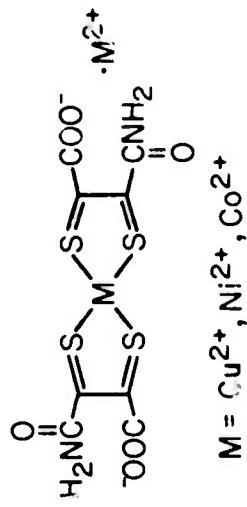
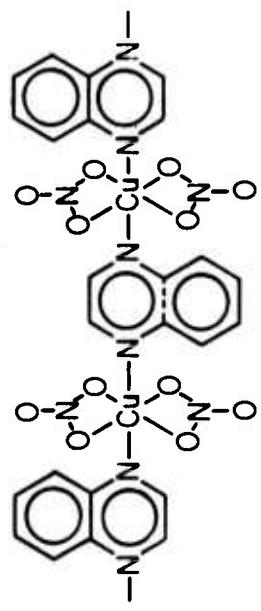
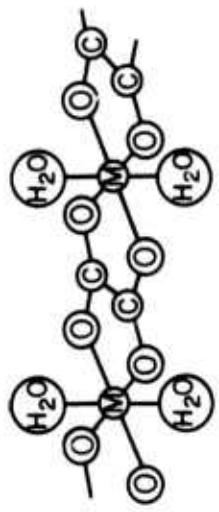
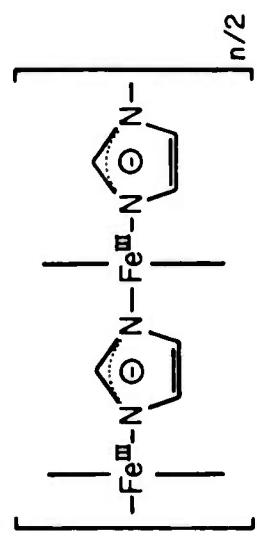


Figure 2. Examples of Structurally 1-d Materials with Multiatom Bridging.

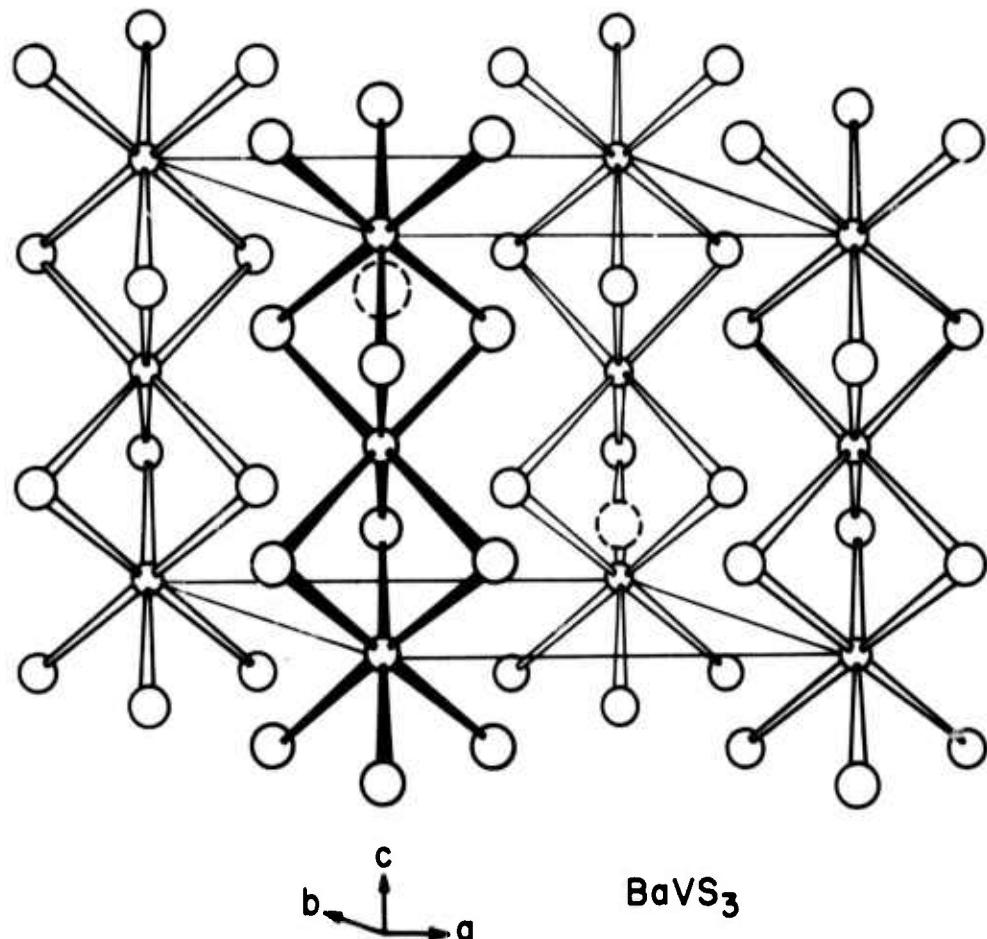


Figure 3. Crystal Structure of BaVS_3 .

hexagonal (V-V distance = 2.80\AA) to orthorhombic. The trigonal component of the crystal field might be expected to result in two bands which are derived from localized metal orbitals of e_{1g} and a_{1g} symmetry. The d^1 electronic configurations of the vanadium atom would give a half filled band if one makes the reasonable assumption that the a_{1g} localized orbitals are lowest in energy. However, since the $S^=$ ion is a weak crystal field ligand and the energy separation between the e_{1g} and a_{2g} orbitals is expected to be small, it is likely that the corresponding bands overlap in a fashion similar to that proposed by Goodenough¹¹ for the axial field split t_{2g} (O_h symmetry) orbitals in rutile so that the band is in fact $1/6$ filled. Further transport and diffraction measurements are needed in order to define the metallic behavior of this system.

There are a few linear chain dibridged octahedral complexes such as those shown in Figure 4 which, although they are not electrically conducting, do show evidence for metal-metal bond formation with alternating long and short metal-metal bonds along an infinite chain. These complexes contain transition metal ions with $4d^1$, $4d^2$ and $5d^1$, $5d^2$ configurations and it is expected that metal-metal bonding for these atoms could take place at a distance $\leq 3.5\text{\AA}$.

These compounds may be examples of materials with a half filled a_1 bond which have undergone a Peierls distortion to an insulating dimerized ground state. It would be of interest to determine if, in fact, one could find a metal-

Some Dibridged Linear
Chain Systems

	M-M	(\AA)
αNbCl_4	3.06	3.76
NbI_4	3.31	4.36
NbOCl_2	3.14	3.56
NbOBr_2	3.12	3.98
NbOI_2	3.16	4.32
MoOCl_2	2.93	3.64

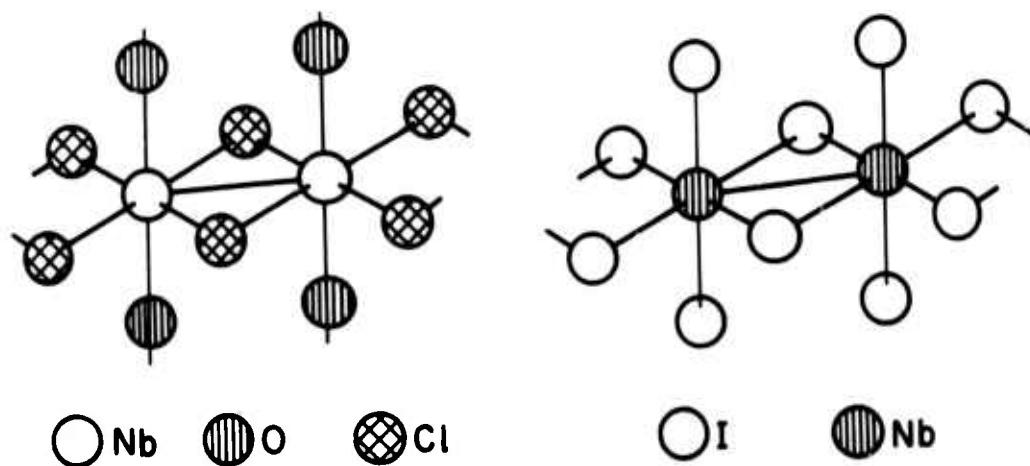


Figure 4. 1-d Chain Structures Formed From Edge Shared Octahedra.

insulator transition for such a compound at higher temperatures below the melting point.

There are also no known monobridged chain systems which are metallic. Mixed valence complexes such as $M^{II} M^{IV} Br_3(NH_3)_3$ where $M = Pd$ or Pt are, however, relevant and the structure of one system is shown in Figure 5. The alternating two long and two short metal ligand bond lengths can again be viewed as the result of the Peierls distortion associated with a half filled band. The metallic state would correspond to a chain with all metal-ligand distances equal, and all the metal atoms in the formal oxidation state of (III). The electronic configuration of each metal atom is then d^7 and the 1-d conduction band can be thought of in the tight binding approximation as being derived from half filled metal d_z^2 orbitals which are directed along the metal chain. At low temperatures, disproportionation would result in the energetically more stable configuration with a repeat distance which is twice that of the undistorted configuration (see Figure 6).

Interrante¹² has investigated the possibility of driving these systems (Figure 5) to a metallic state by the application of high pressures and some results are shown in Figures 7 and 8. For a pressure change of 50 kilobars, one observes a change of more than 0.20\AA° in the metal-metal distance and a corresponding change in bulk conductivity of from 10^{-5} to more than $1 \text{ ohm}^{-1} \text{ cm}^{-1}$. Although Interrante has argued on the basis of bond length changes that the higher conducting phase is not a Pd^{III}

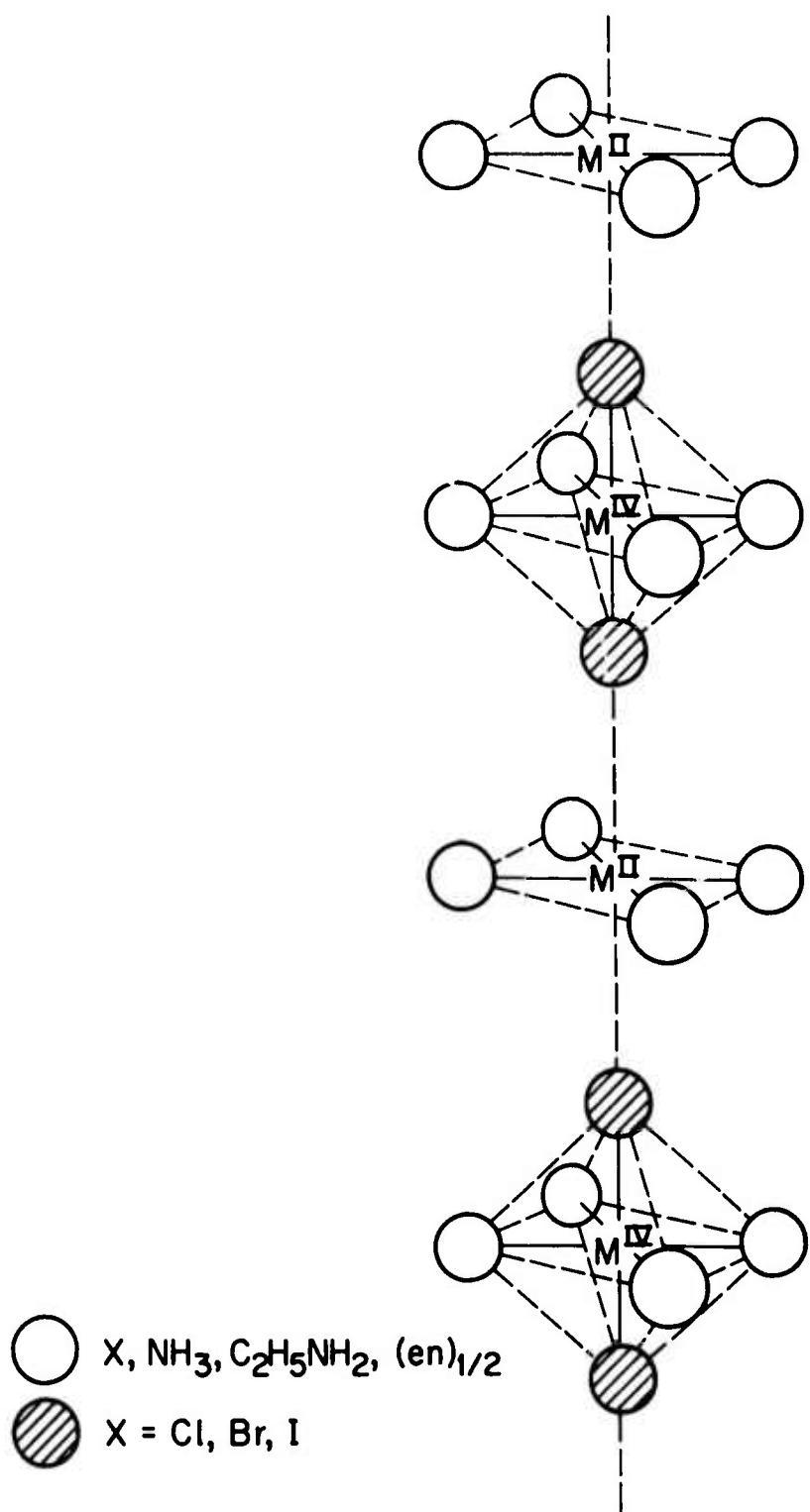


Figure 5. One Dimensional Ligand Bridged Complexes MBX .
 The configurations about each metal atom in the
 plane \perp to the stacking direction are identical.
 (See Figure 1.)

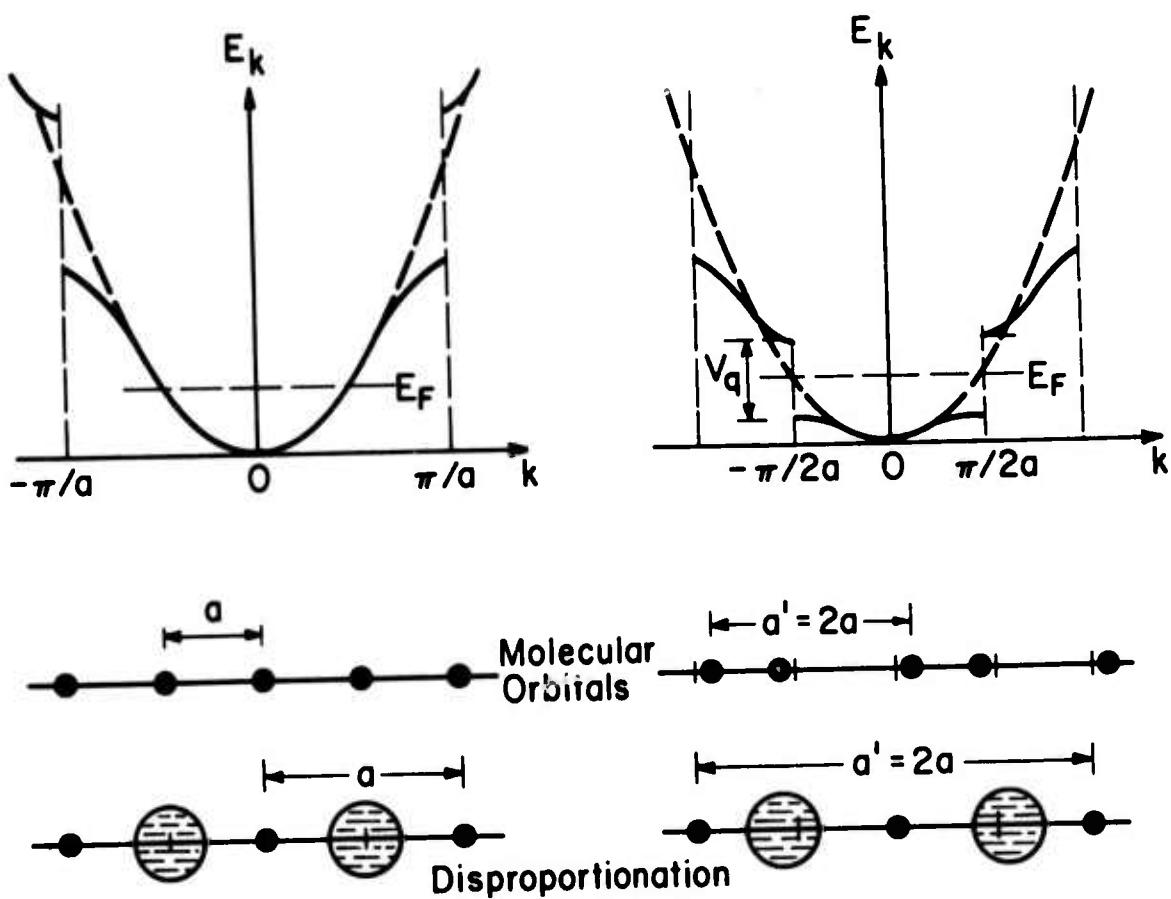


Figure 6. Peierls Distortion Associated with a Mixed Valence Complex $-X-M^{n+}-X-M^{n+}-$. The Peierls Distorted State Corresponds to $\dots X-M^{n+2}X\dots M^{n-2}\dots X\dots$.

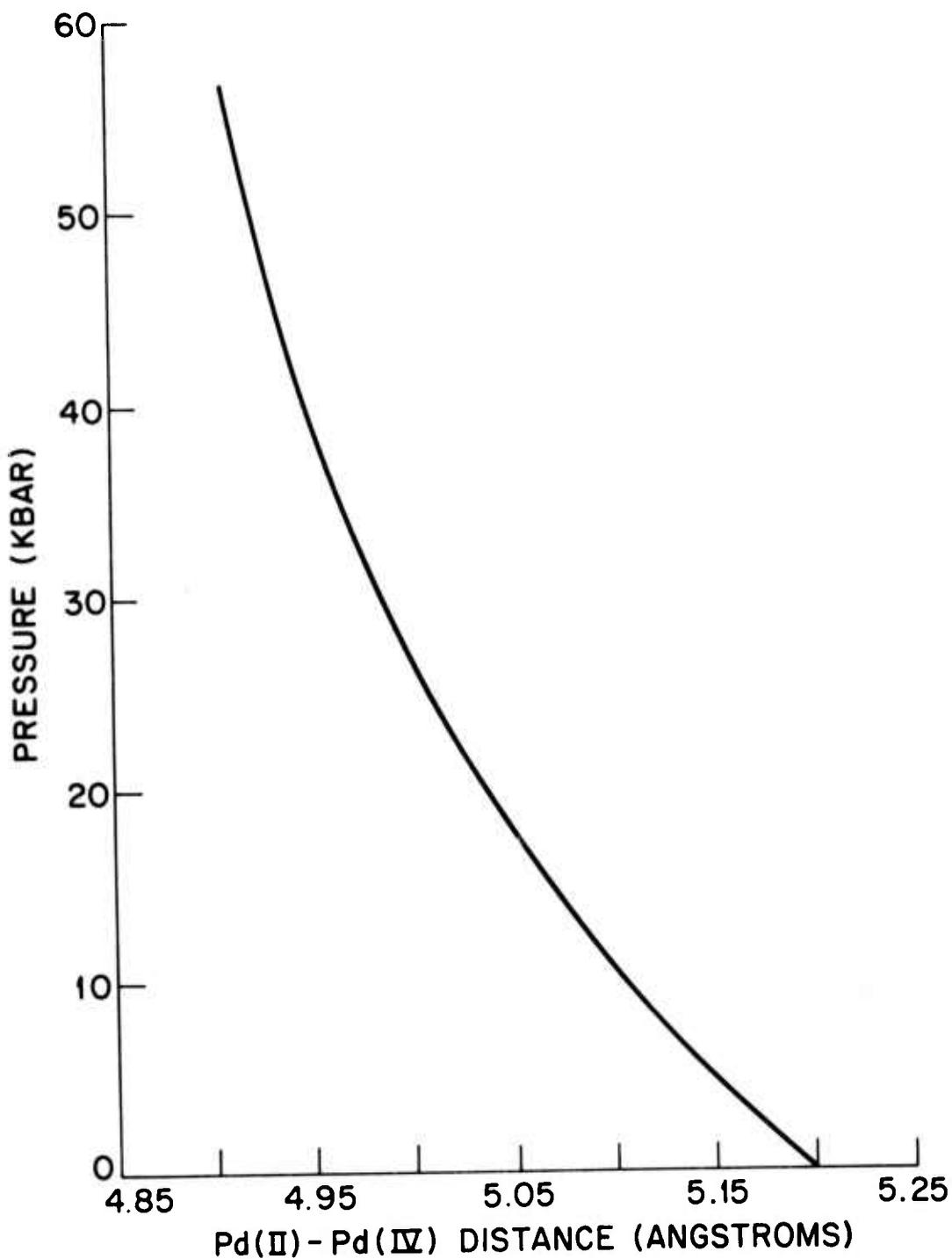


Figure 7. Variation of Pd(II)-Pd(IV) Distance in $\text{Pd}(\text{NH}_3)_2\text{Br}_3$ with Pressure.

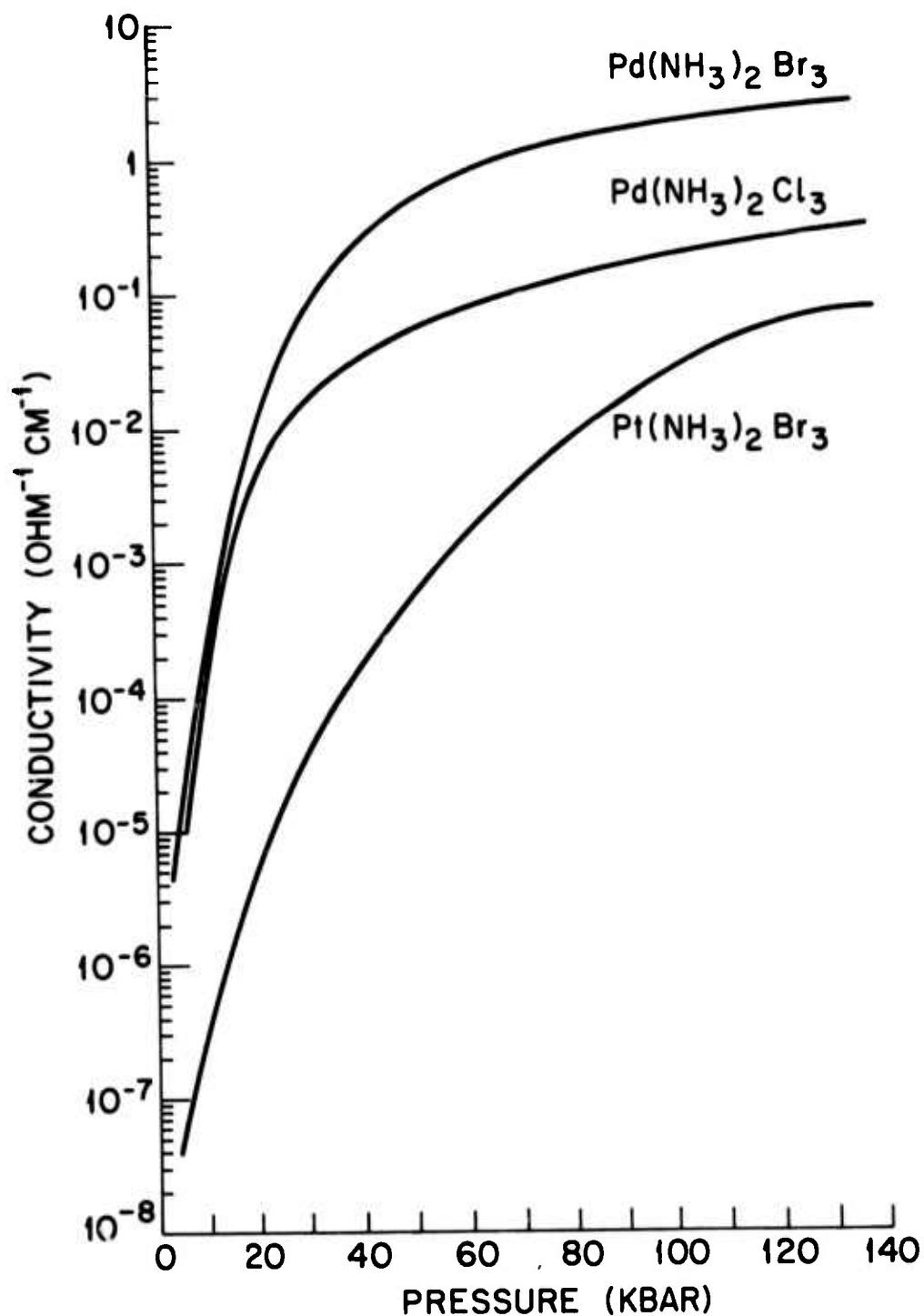


Figure 8. Variation of Conductivity with Pressure for $\text{Pd}(\text{NH}_3)_2\text{Br}_3$.

complex, the conduction band is made up of localized orbitals which are antibonding with respect to the metal-metal interaction. The M-M bond lengths that one would expect for the $\text{Pd}^{(\text{III})}-\text{Br}-\text{Pd}^{(\text{III})}$ arrangement, would not, therefore, be expected to be the average of the $\text{Pd}^{(\text{II})}-\text{Br}-\text{Pd}^{(\text{IV})}$ bond lengths but closer to twice the $\text{Pd}^{(\text{II})}-\text{Br}$ distance. Unfortunately, the experimental conditions do not permit the study of single crystals.

B. Square Planar Transition Metal (d^8) Compounds

The only class of 1-d, inorganic complexes which show metallic conductivity are those derived from square planar d^8 configurations with direct metal-metal interactions as illustrated in Figure 9. However, only the partially oxidized, non-stoichiometric complex of the type given in Table I exhibit metallic behavior. Partially oxidized d^8 complexes also can be made with rhodium and iridium. In all cases, partial oxidation of the d^8 complexes is necessary because the highest energy band in the metal stacked d^8 configuration is a filled band made up of d_z^2 localized metal orbitals. The room temperature conductivities of these compounds are on the order of 10 to $10^2 \text{ ohm}^{-1} \text{ cm}^{-1}$. The structures of all of the d^8 one dimensional systems are characterized by short metal-metal distances of $2.85 \pm 0.1 \text{ \AA}$. In addition, the ligands on adjacent metal atoms are rotated in such a way as to minimize non-bonded interactions. The structure and physical properties of $\text{K}_2\text{Pt}(\text{CN})_4\text{Br}_{0.3}^{\circ}\text{H}_2\text{O}$ have been studied the most extensively^{13, 14} and represent the first

TABLE I. Examples of Partially Oxidized d^8
Linear Chain Complexes

	Oxidation State of the Platinum	Pt-Pt Distance
$K_2 [Pt(CN)_4]Cl_{0.32} \cdot 3H_2O$	+2.32	2.880\AA°
$K_2 [Pt(CN)_4]Br_{0.30} \cdot 3H_2O^*$	+2.30	2.887
$Mg [Pt(CN)_4]Cl_{0.28} \cdot 7H_2O$	+2.28	2.985
$K_{1.74} [Pt(CN)_4] \cdot 1.8H_2O$	+2.26	2.96
$H_{1.60} [Pt(C_2O_4)_2] \cdot 2H_2O$	+2.40	2.80
$Li_{1.64} [Pt(C_2O_4)_2] \cdot 6H_2O$	+2.36	2.81
$Ba_{0.84} [Pt(C_2O_4)_2] \cdot 4H_2O$	+2.32	2.81

*also known as Krogmans Salt or KCP.

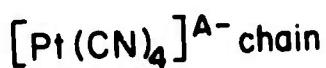
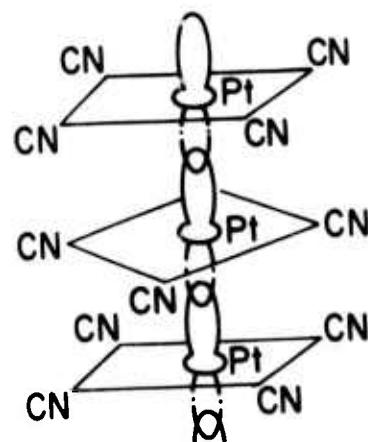
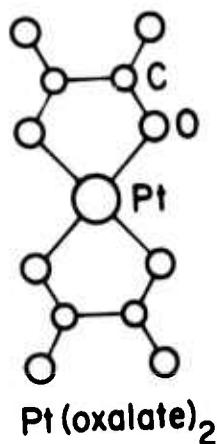
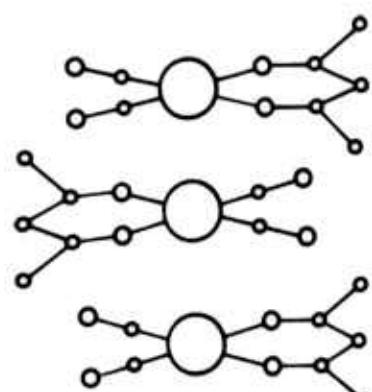
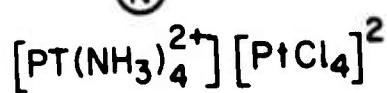
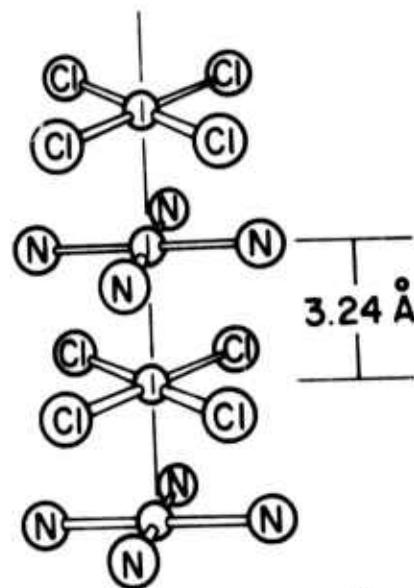
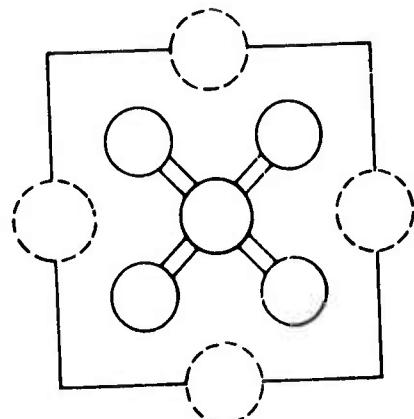


Figure 9. Examples of d^8 (Square Planar) Metal Stacked Complexes.

documented example of a highly 1-d metallic conductor.

Although, as pointed out earlier, the compound $K_2Pt(CN)_4Br_{0.3} \cdot 3H_2O$ (KCP) was first isolated in 1842, the identification of it as a metallic material was not confirmed until 1969¹⁴ when Krogman was able to show that the structure of the $Pt(CN)_4^{n-}$ portion of the chain was that shown in Figure 9. The mechanical properties of KCP resemble those of most inorganic salts with the added complication that the water molecules are readily lost even at 300°K with a drastic change in the physical properties, including a loss of conductivity.¹⁴⁻¹⁷ Therefore, physical studies above 300°K have not been pursued. A second difficulty is that KCP is also an electrolytic conductor, and special techniques must be used in forming electrical contacts with the crystal surface.^{18, 19} Elastic neutron diffraction studies by Williams have shown that the original planar configuration for the proposed $Pt(CN)_4^{2-}$ anion by Krogman is in fact incorrect and the crystallographic space group is non-centrosymmetric (Figure 10)²⁰. The water molecules are hydrogen-bonded to the cyanide nitrogen and the bromine atoms in two ways and the bromine atom sites are apparently randomly occupied by both bromine atoms and water molecules. Unlike the organic charge transfer conductor, TTF-TCNQ, the conductivity of KCP decreases smoothly with decreasing temperature¹⁵. Many other features of KCP, however, are similar to those proposed later for TTF-TCNQ such as the presence of a Peierls distortion and the pinning of a one-dimensional charge density wave.¹³ The

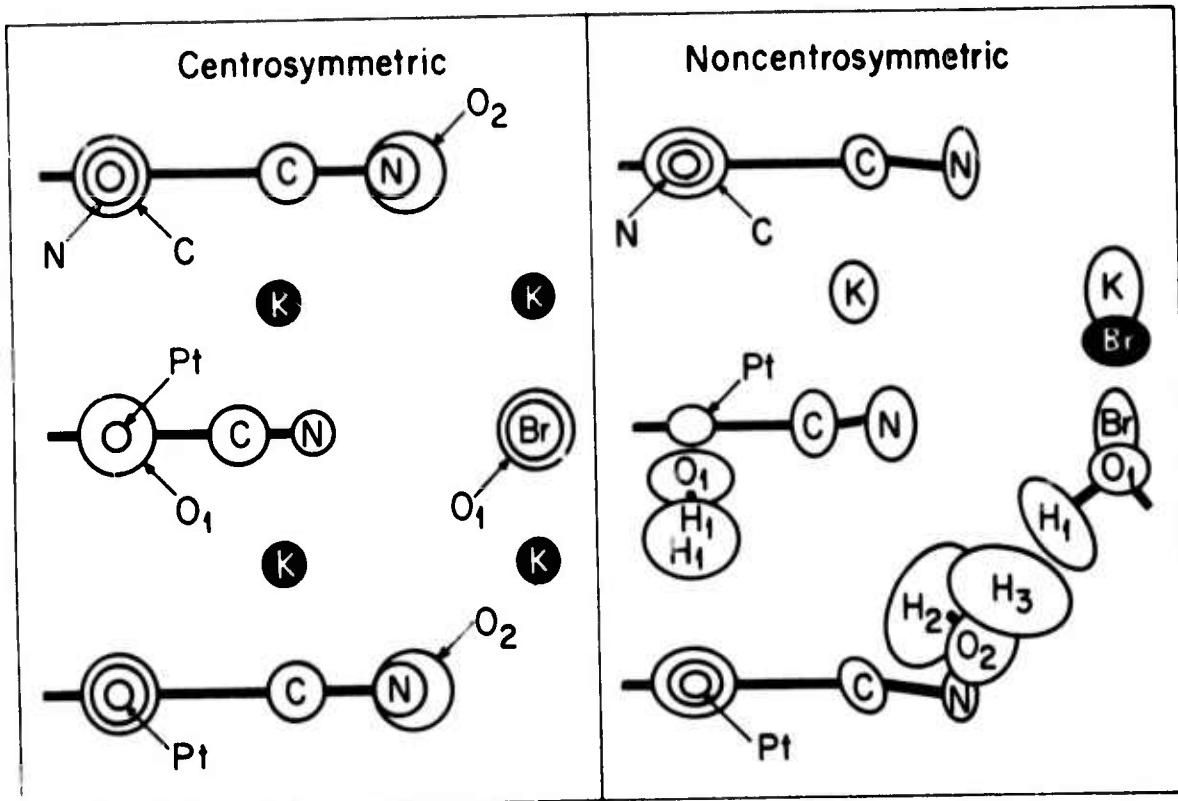


Figure 10. Structure of $K_2Pt(CN)_4Br_{0.3} \cdot 3.2H_2O$ as determined by early X-ray and later neutron diffraction experiments.

diffuse X-ray and neutron scattering from KCP studied by Comes and coworkers^{21, 22} as a function of temperature indicated a gradual transition from a one- to a three dimensional ordered state. Furthermore, the presence of a soft phonon mode observed by neutron inelastic scattering²³ at 300°K implies that the charge density wave (Peierls) distortion has not taken place. More recent studies²⁴ suggest instead that the inelastic scattering in the vicinity of the $2k_F$ instability falls to $E \approx 0$ for all temperatures in the range 78-300°K. These results are interpreted to indicate that at 300°K the CDW distortion has already manifested itself in the lattice and that the development of interchain correlations at low temperatures, which is observed by both neutron and X-ray elastic scattering, is not due to a Peierls-type transition. The picture is then one of a sliding (quasi-static) charge density wave in the parallel $\text{Pt}(\text{CN})_4$ chains with a temperature dependent interchain correlation length which never exceeds a few chain spacings. At higher temperatures, the elastic scattering shows diffuse streaks which can be attributed to uncorrelated quasistatic charge density waves. The intra-chain correlation length was estimated to be $\sim 300\text{\AA}$ from these experiments. In short, in this picture, the conductivity behavior of KCP from 300°K to 4.5°K is similar to that suggested by Heeger and coworkers (see next section) below 53°K for TTF-TCNQ. Other types of distortions in oxidized d^8 systems have been suggested by Krogman²⁵ from elastic scattering experiments as indicated in Figure 11. This is not surprising since the

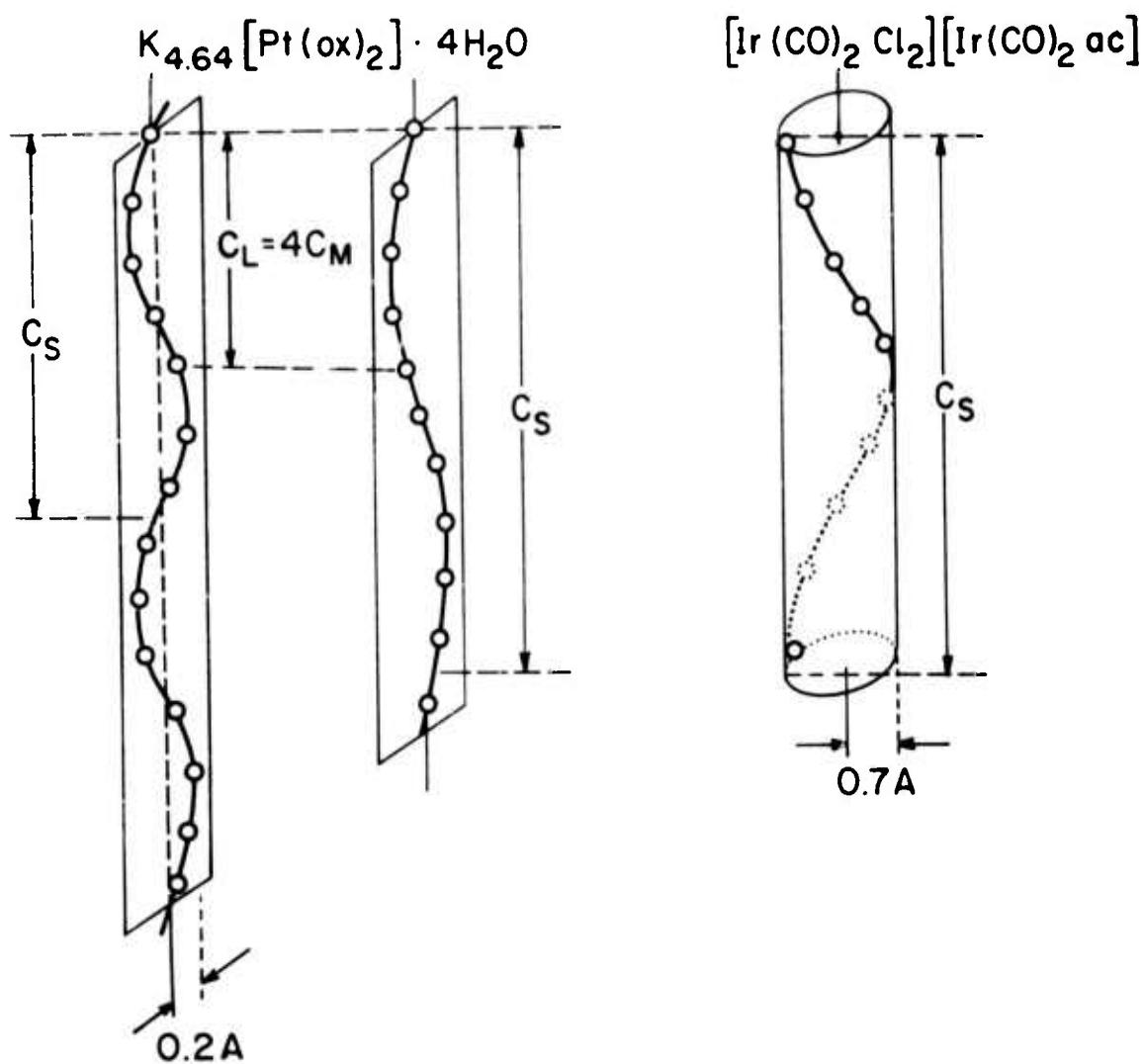
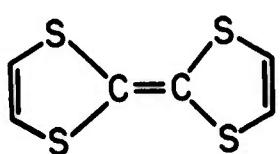
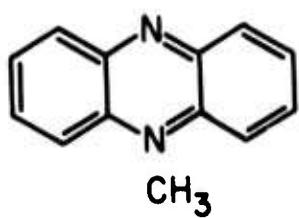


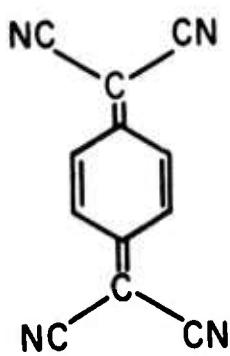
Figure 11. Transverse Sinusoidal and Helical Distortions Reported for Partially Oxidized d^8 Complexes.³⁰



TTF



NMP



TCNQ

Figure 12. Molecular Structures of TTF, NMP and TCNQ.

symmetries of these distortions must depend on the specific nature of interchain interactions which can be expected to vary for each material.

C. Organic Charge Transfer Salts

The criteria for designing organic systems which exhibit metallic behavior, as summarized by workers at The University of Pennsylvania^{26, 27}, are

(1) the existence of unpaired electrons within the molecular units (charge transfer);

(2) a uniform crystal structure consisting of linear, parallel stacked columns of flat planar molecules with significant intermolecular overlap leading to band widths such that, in the absence of electron-electron repulsions, the associated electronic band structure would be metallic;

(3) molecular units exhibiting intramolecular electron correlation to allow charge density to reside on diametrically opposed parts of the molecule; and

(4) highly polarizable molecular units having minimal size to diminish costly electron-electron repulsive interactions.

A similar set of criteria are given by the workers at John Hopkins University.²⁸.

Donor molecules such as N-Methyl phenazine (NMP) and tetrathifulualene (TTF) satisfy these requirements (Figure 12). By far, the most highly conducting materials have been obtained with TCNQ as the acceptor unit. The stacking of the TCNQ molecules in the metallic charge transfer complexes is illustrated

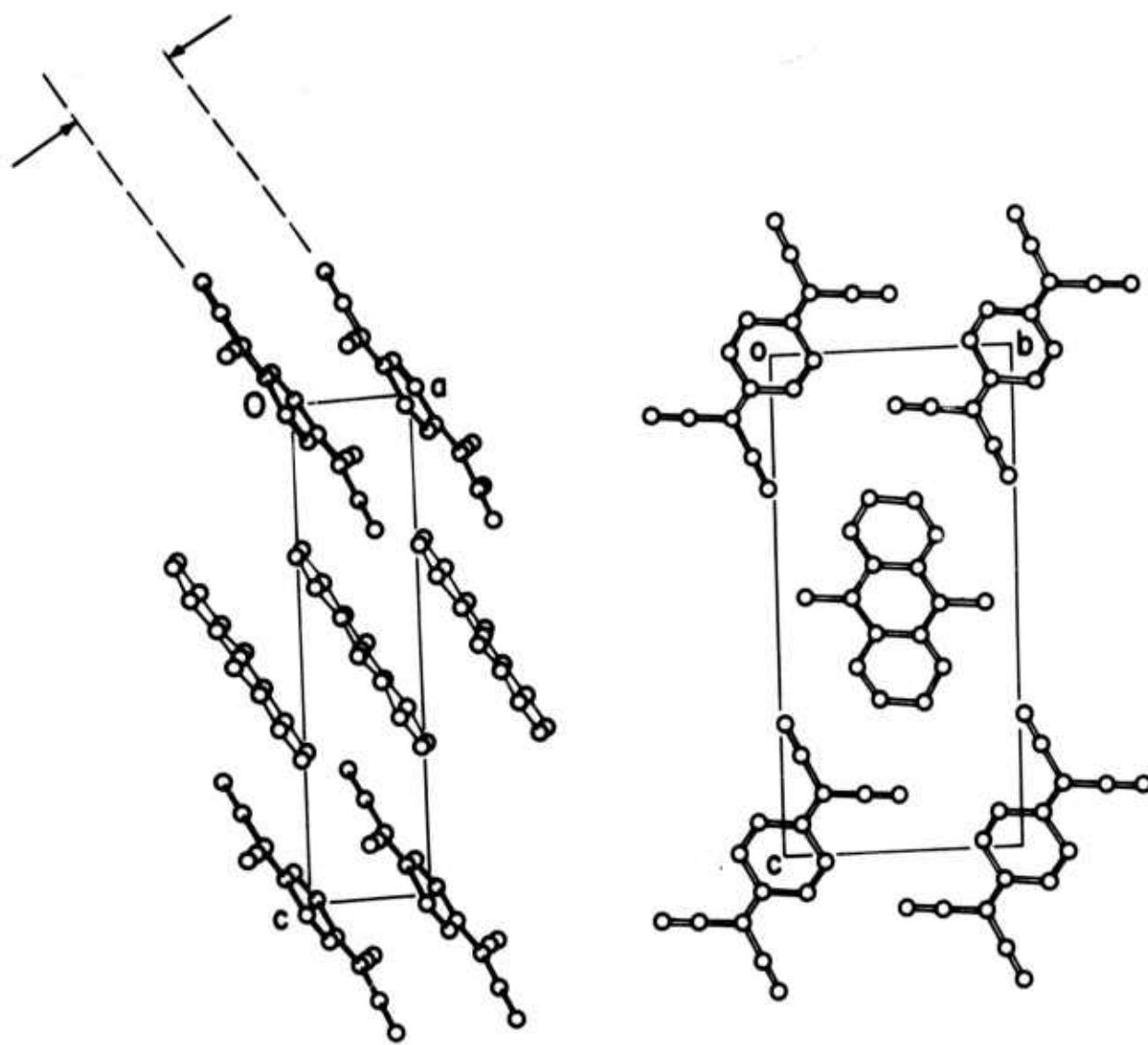


Figure 13. Crystal Structure of NMP NCNO. The NMP molecule is disordered and both positions of the Methyl Group are shown.

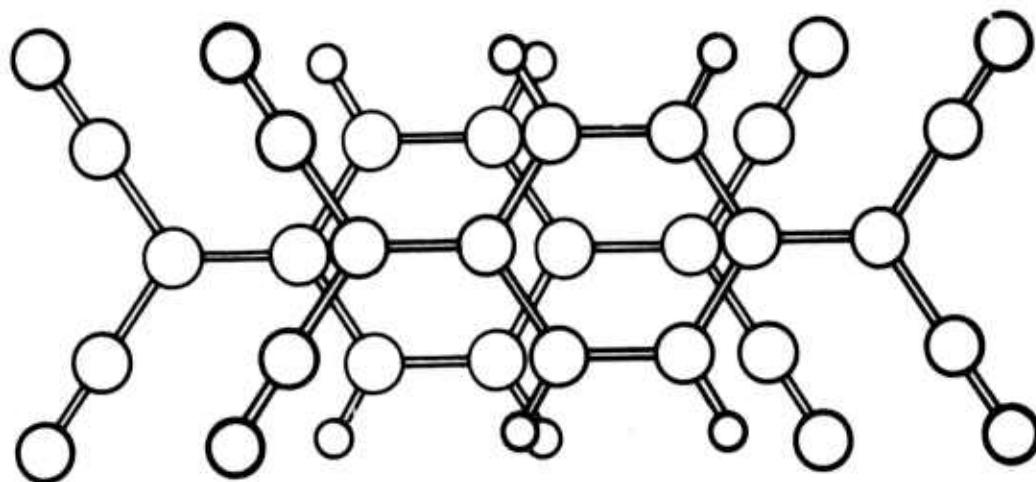


Figure 14. Stacking of TCNQ Molecules in NMP TCNQ, TTF-TCNQ and Other Metallic TCNQ Systems.

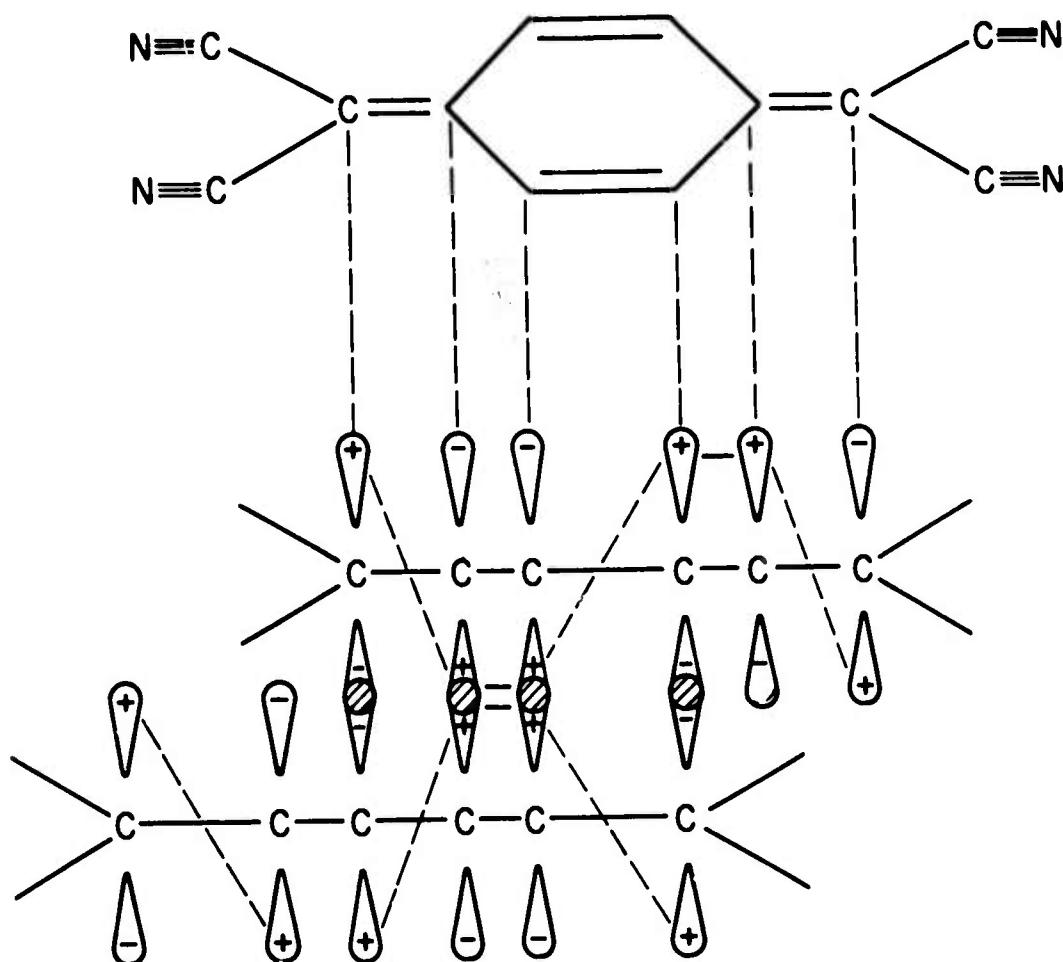
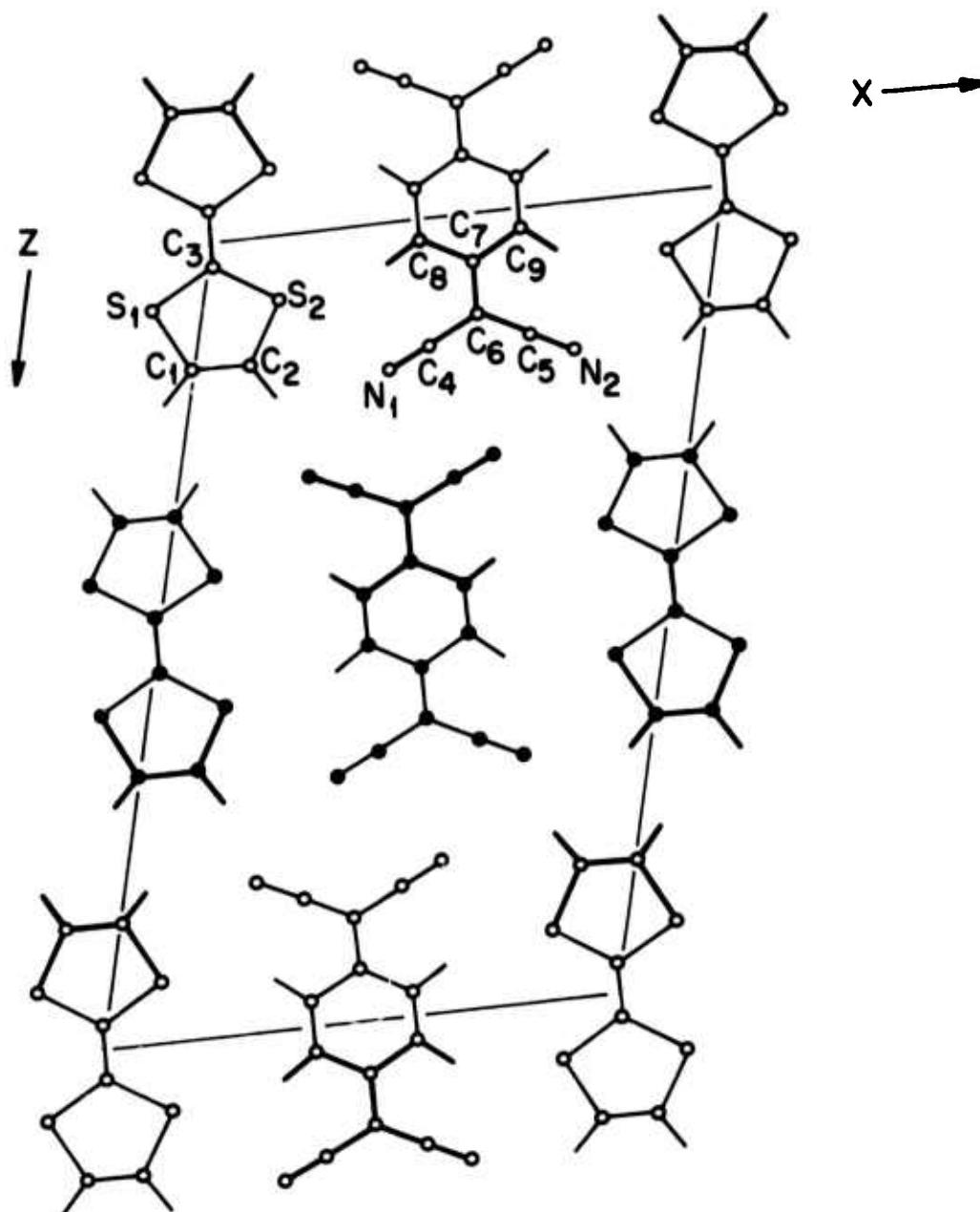


Figure 15. Phasing of Molecular Orbitals on Adjacent TCNQ Molecules.



TTF TCNQ

Figure 16. Crystal Structure of TTF-TCNQ.

in Figure 13 for NMP-TCNQ which has a single crystal conductivity along the stacking axis of $>10^2 \text{ ohm}^{-1}\text{cm}^{-1}$ at room temperature. Each TCNQ molecule is displaced relating to the one adjacent to it as shown in Figure 14. As a result the molecules are stacked with the planes of the TCNQ molecules inclined at an angle of $\sim 30^\circ$ with respect to the stacking direction.⁴ The fact that the planes are not stacked directly over each other permits the proper phasing of localized molecular orbitals to give a large transfer integral between molecules bringing electron deficient and electron rich regions of the adjacent molecules into near proximity of each other. This minimizes coulombic charge separation as shown in Figure 15.²⁶

The TCNQ interplanar spacing in this compound is 3.26°\AA which is at the upper end of the TCNQ-TCNQ distances found in conducting TCNQ materials. The crystal structure of the related charge transfer compound, TTF-TCNQ²⁹ is shown in Figure 16. As described for the TCNQ molecule in NMP-TCNQ above, the planes of both the donor and acceptor molecules are inclined with respect to the stacking axis (y direction) with a TCNQ-TCNQ separation of $3.168(2)^\circ\text{\AA}$. The former distance is one of the shortest known for an interplanar spacing between carbon atoms. (c.f. graphite at 3.34°\AA .)

TTF-TCNQ has been the most extensively studied member of the organic charge transfer class of compounds. A wide range of experiments have examined various electronic, structural, thermodynamic and transport properties. The measurement of the tempera-

ture dependence of the conductivity $\sigma(T)$ on the first synthesized crystalline samples of TTF-TCNQ indicated a very large conductivity at 300°K ($\sigma \approx 5 \times 10^2 (\Omega\text{cm})^{-1}$).³⁰ This value was the largest reported for an organic compound at that time. The temperature dependence was unusual in that σ showed a sharp maximum of 10^4 $(\Omega\text{cm})^{-1}$ at $T \approx 60^\circ\text{K}$.³⁰ In addition, the crystalline conductivity was found to be highly anisotropic which is apparent from the crystal structure mentioned above with segregated TTF and TCNQ stacks along the b axis of the crystal.

Independent syntheses and measurements of $\sigma(T)$ ³¹ indicated a few carefully prepared samples to have larger peak values approaching $\sigma_{\text{max}} \approx 10^5 (\Omega\text{cm})^{-1}$ at 58°K . Further examination^{32, 33} of these remarkable results, including studies of the DC probe techniques in highly anisotropic compounds, methods of chemical synthesis, examination of crystal purity, and methods of purification have left a few remaining discrepancies³⁴ in the actual value of σ_{max} . However, there is general agreement that the value of σ_{max} is bounded and sensitive to crystalline perfection and purity. Most of the results from several laboratories are now in agreement to within a factor of 2.

The importance of surface effects in the measurement of conductivities is brought out again in a recent work which reports the difference in the charge transfer between surface and bulk species in TTF-TCNQ.³⁵ According to these angle reflection dependent X-ray photoelectron (ESCA) studies of TTF-TCNQ on the N_{1s} atom, a discrete layer (1-5 layers of complex) occurs on the

surface where no charge transfer takes place relative to the 0.8 ± 0.2 electron transfer in the bulk sample. Furthermore, the better conductors show the largest discrepancy between surface and bulk charge transfer effects.³⁵

The frequency dependence of the conductivity, $\sigma(\omega)$, of TTF-TCNQ has also been studied in a range extending from the microwave through the near infrared. Results at 10 GHz agree substantially with the DC values.^{36,37} Heeger, Garito and co-workers have interpreted^{38,39} the infrared results obtained from absorption and reflection measurements in terms of single-particle interband transitions corresponding to a band gap of $\hbar\omega_g = 0.14$ eV associated with a Fröhlich-Peierls state. At higher frequencies $\sigma(\omega)$ is Drude-like with an associated plasma frequency of $\hbar\omega_p = 1.2$ eV. Phonon modes corresponding to molecular vibrations are superposed on this structure.

The maximum in the conductivity at 58°K is suggestive of a phase transition. Definite evidence for a phase transition was observed by measurements of the temperature dependence of the specific heat.⁴⁰ Other evidence for a phase transition includes measurements of the thermoelectric power⁴¹, the thermal reflectance⁴², the noise spectrum⁴³, the thermal conductivity⁴⁴, and the longitudinal (parallel to b) and transverse conductivity ratio $\sigma_{\parallel}^b(T)/\sigma_{\perp}^a(T)$.^{45,46} Of these, the latter two measurements also show an additional phase transition near $T = 38^\circ\text{K}$. The temperature dependence of the susceptibility also shows evidence of a phase transition.⁴⁷ The susceptibility decreases

smoothly with decreasing temperature above 55°K. Below this temperature it falls more rapidly to the molecular TTF-TCNQ value calculated from Pascal's rules. The number of unpaired electrons existing above 55°K diminishes rapidly at lower temperatures and approaches zero.

The nature of the phase transition is important in interpreting the conductivity results. Low temperatures X-ray scattering work^{4,8} indicates the existence of an incommensurate one-dimensional super-lattice below $T = 55^{\circ}\text{K}$ with a relatively large coherence length. These observations are consistent with a metal-semiconductor Peierls-type transition.⁷

This model is further substantiated by the results of neutron scattering. Mook⁹ has observed a giant Kohn anomaly (soft mode) in the longitudinal acoustic phonon branch at a wave number that is consistent with the results of the X-ray data.

A number of mechanisms have been proposed to account for the conductivity and its temperature dependence in TTF-TCNQ. Of these, the model of Heeger and Garito^{4,9}, based on the picture proposed by Lee, Rice and Anderson^{5,0} (to be denoted as the charge density wave (CDW) model), has been worked out in greatest detail to explain the existing data resulting from a variety of experiments. This model assumes, on the basis of experimental arguments following from NMR and susceptibility measurements, that the electron-phonon interaction is in the intermediate coupling range and far stronger than electron-electron interactions. The phase transition at 58°K is identified as a Fröhlich-Peierls

transition. At lower temperatures the system is a semiconductor characterized by a very large DC dielectric constant parallel to the b axis ($\epsilon_{\parallel}'' \sim 3500$) and a more modest value ($\epsilon_{\perp}'' \sim 5$) perpendicular to that axis.^{46, 51} Above 58°K the system becomes a highly anisotropic one-dimensional metal in which the electrons are localized on the TCNQ chains (and the holes, presumably, at the TTF). There are two kinds of electronic excitations in such a system; a collective mode (or charge density wave) of the coupled electron-phonon system, and single particle excitations across the pseudogap. The relatively large values of $\sigma(\omega)$ at DC and in the microwave range are associated with the collective excitation; the peak near 1000 cm⁻¹ in the infrared corresponds to the single particle excitations. The rise in $\sigma(T)$ as the temperature is lowered towards the phase transition is attributed to the onset of electron conduction in the charge density wave state and the decrease below 58°K to the fact that the collective mode becomes increasingly pinned. The 38°K transition is associated with complete pinning. It should be pointed out, however, that to our knowledge no theoretical calculation of $\sigma(T)$ based on the CDW model has yet been done.

A number of mechanisms have been proposed that explain the observed values of σ_{\max} in terms of theories not involving a collective mode.^{34, 52-54, 54a}

The single particle point of view is supported by the interpretation of the temperature dependence of the thermal conductivity⁴⁰, which suggests the collective mode contribution to

to be negligible. The question as to whether the value of σ_{\max} is sufficiently large to suggest an explanation in terms of a collective mode, has also been raised.³⁴ The charge density wave (CDW) model attributes the observed structure in $\sigma(\omega)$ entirely to electrons in the TCNQ chains and assumes implicitly that the contributions to $\sigma(\omega)$ due to the TTF are negligible. This leaves open the question as to the role of the TTF chains in the conduction process. Indeed, the conductivity along the TTF chain in the mixed valence compound TTF-Br_{0.74} is quite high at room temperature ($\sigma=300(\Omega\text{cm})^{-1}$). If the TTF conductivity is appreciable, it may be necessary to consider an energy band structure that involves both chains, and also possible interchain electron-electron interactions^{54b} as hybridization between them. Band calculations^{55, 56} have been carried out using LCAO-CNDO calculations that show that the conductor band associated with the TTF and TCNQ stacks respectively cross at the Fermi level and hybridize then as a result of interchain coupling. The resulting band structure is semi-metallic and the density of states in the hybridization region is small, as is its width ($\sim 400^\circ\text{K}$). This width is roughly consistent with the gap of 22 meV found from photoconductivity measurements⁵⁷, and with a mean-field transition temperature near 55-60°K in the vicinity of the observed anomalies. The fact that the conduction along TTF chains may be appreciable is supported by measurement of thermoelectric power.⁴¹ The observed change in sign can be interpreted as suggesting low and high temperature conductivities along the

TTF and hybridized TTF-TCNQ chains respectively. The g -values obtained from EPR measurements, indicate that the states contributing to the susceptibility are fully hybridized above the transition temperature and primarily TTF states below.^{58,59}

If, on the other hand, it is supposed, as in the CDW model, that σ along the TTF chain is sufficiently small that the electronic transport is truly diffusive, a band theoretical description of the TTF electronic states is open to question.

The view that the electron-phonon interaction is strong in the TTF-TCNQ system appears to be shared by many of the groups contributing actively to this field. There appears also to be general agreement that there is a phase transition at 58°K which is variously termed as a "metal-insulator" or "Fröhlich-Peierls" transition depending on the theoretical viewpoint. Whatever picture one adopts, however, it is difficult to see how the sharp peak in $\sigma(T)$ is to be explained without involving some kind of resonance between the electron and phonon systems.

It seems important to explore the behavior of $\sigma(\omega)$ in the microwave region both above and below the transition temperature in more detail. Information concerning the shape and shift with temperatures of the peak in $\sigma(\omega)$ in this frequency range, associated by the Pennsylvania group with the collective mode, is essential to a resolution of the discussion concerning theoretical models that remains at the present time.

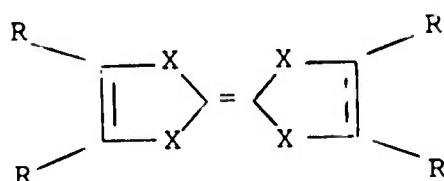
Several chemical derivatives of TTF have been developed by various groups. The focus on the IBM research has been to

examine changes in electronic properties for structural derivatives of TTF-TCNQ. It has been determined that TSF (the selenium analogs of TTF) has a lower transition than TTF and a somewhat lower value of $\sigma_{\text{max}}/\sigma_{\text{RT}}$. "Alloys" of TTF and TSF can be made and used to study the effect of substitution in the donor chain in TTF-TCNQ on the electronic behavior as a function of concentration.

Research at John Hopkins University has been directed towards the modification of TTF-TCNQ by use of hetero-atoms and substituent groups to optimize bandwidths, band overlap, charge transfer and interchain coupling. The compounds which have been synthesized and their properties are listed in Table II. The compound HMTSF (Figure 17) is unique among known charge transfer complexes in that it exhibits metallic behavior to $<1.1^{\circ}\text{K}$.⁶⁰ No evidence of superconducting behavior has been reported. Another new organic electron donor tetramethyltetraselenofulvalene (TMTSF) forms a 1:1 charge transfer salt with TCNQ which occurs in two crystal forms. The first is a red insulator whose structure consists of chains of alternating donor and acceptor molecules; the second is a black quasi-one-dimensional conductor in which donor and acceptor are separately stacked. With cooling, the conductivity of the black form rises from its 300°K value of $\sim 1200\Omega^{-1}\text{cm}^{-1}$ to a maximum of $\sim 8000\Omega^{-1}\text{cm}^{-1}$ near 65°K , where a metal-to-insulator transition occurs.

HMTSF and TMTSF both also form conducting salts with another electron acceptor, TNAP. Preliminary dc and microwave measurements indicate that both HMTSF-TNAP and TMTSF-TNAP remain

TABLE II. Comparison of Physical Properties of Derivitized
TTF TCNQ Complexes. In all cases the Acceptor
Molecule is TCNQ.



	R	X	σ_{RT}	σ_{max}/σ_{RT}	T_{max} (°K)
TTF	H	S	400	20	59
ATTF	2H, 2CH ₃	S	50	25	~50
TMTTF	4CH ₃	S	1000	1	-
TMTTF	4CH ₃	S	350	15	60
HMTTF	-CH ₂ -CH ₂ -CH ₂ -	S	500	3-5	~80
TSF	H	Se	800	12	40
DTDSF	H	2S, 2Se	500	7	64
TMTSF	CH ₃	Se	1000	5	71
HMTSF	-CH ₂ -CH ₂ -CH ₂ -	Se	1800-2000	No Transition	

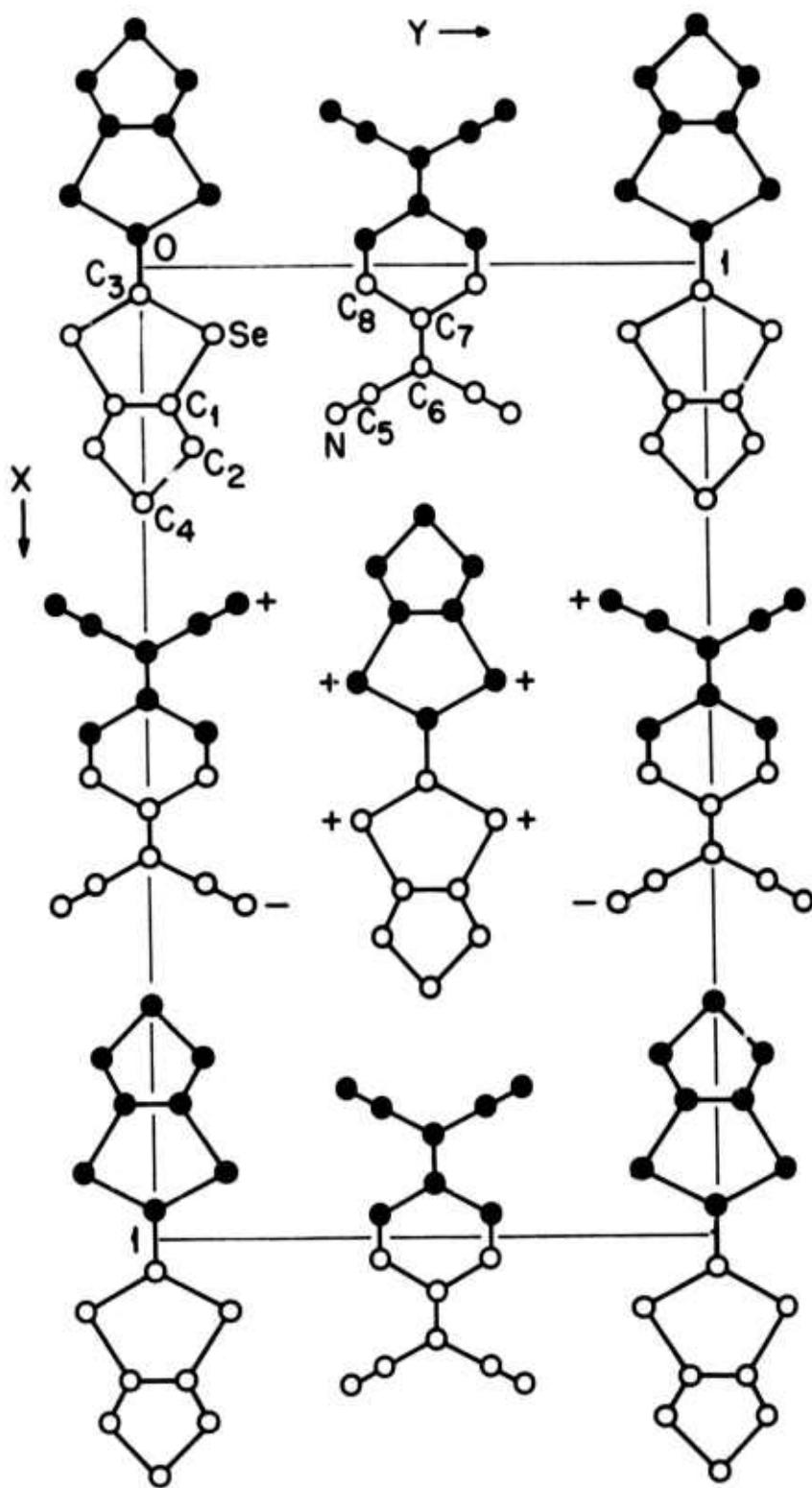


Figure 17. Crystal Structure of Hexamethylene Tetra Seleno fulvalene (HMTSF).

metallic to 11°K and display resistance minima at somewhat high temperatures than HMTSF-TCNQ.

Using the stated general guidelines for the design and synthesis of organic metals, the Pennsylvania group has been studying the effects on electronic and structural properties of TTF-TCNQ related salts introduced by systematic changes in:

- i) the ionization potential and polarizability of TTF related donors by R group substitution and diselenium substitution;
- ii) interchain coupling and steric hindrance in TTF related donors;
- iii) electron affinity of substituted TCNQ acceptors; and
- iv) disorder introduced by asymmetric substitution of R groups, or diselenium substitution on TTF related donors as well as by mixed alloys such as $\text{TTF}(\text{TCNQ})_x(\text{TCNQF}_4)_{1-x}$ and $(\text{TTF})_x(\text{DSeDTF})_{1-x}$ TCNQ.

The compaction resistivities of a large number of organic charge transfer complexes has been recently given by DuPont workers.³⁵ The results are reproduced here in Table III with the compounds being identified on the accompanying diagram. The A/B ratios are the relative magnitudes of the N_{1s} ESCA spectra for the surface and bulk compounds as described previously.

Additional work should be encouraged in designing and synthesizing new organic charge transfer compounds with an attempt to systematically catalog the molecular dependence on the solid state properties. Again we emphasize the importance

TABLE III. Resistivities* and A/B Ratios for Various Complexes**

Compound	$\log_{10} \rho$ (ohm-cm)	A/B
1. TTF-TCNQ	-1.2	1.16
TTF-TCNQ	-1.2	0.83
2. TTF-TCNQBr	-1.0	1.00
3. TTF-TCNQBrCH ₃	-1.0	0.94
4. TTF-TCNQEt ₂	-1.0	0.95
5. NMP ⁺ TCNQ ⁻	-1.0	1.11
6. TTF-TCNQCl	-0.7	1.00
7. TTF-TCNQClCH ₃	-0.5	1.10
8. TTF-TCNQBr ₂	-0.4	1.18
9. TTN-TCNQ	-0.2	0.94
TTN-TCNQ	+0.5	0.91
10. Tl ⁺ TCNQ ⁻	+1.8	0.45
11. TTF-TCNQCl ₂	+2.9	0.90
12. TTF-TCNQF ₄	+2.9	0.54
13. K ⁺ TCNQ ⁻	+3.7	0.42
14. Cs ⁺ TCNQ ⁻	+4.5	0.40
15. Na ⁺ TCNQ ⁻	+4.5	0.30
16. TMDSA-TCNQF ₄	+5.0	0.40
17. Li ⁺ TCNQ ⁻	+5.3	0.30
18. S-Se-TCNQ	+5.6	0.21
19. S-Se-TCNQF ₄	+5.6	0.44
20. Ag ⁺ TCNQ ⁻	+5.9	0.30
21. TTF-DDQ ⁺⁺⁺	+6.8	0.29
22. TMDSA-TCNQCl ₂	+7.0	0.26
23. TMDSA-TCNQ	+9.3	0.18
24. (O ₃ PCH ₃) ⁺ TCNQ ⁻	+10.0	0.40

* R. C. Wheland and J. L. Gillson, to be published.

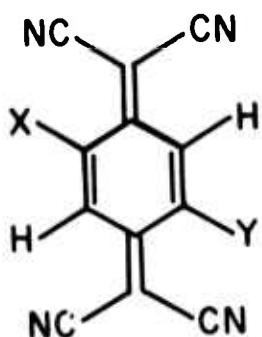
** All data taken on powdered samples.

† A/B uncertain due to overlap of N_{1s} from donor and acceptor.

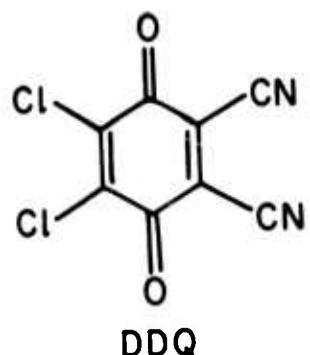
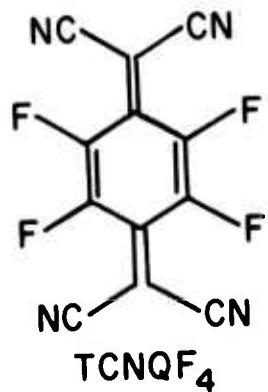
†† A/B uncertain due to partial N_{1s}-Tl_{4d}_{3/2} overlap.

††† A/B uncertain due to small, if any, ΔE_B between N_{1s} main line and satellite structure.

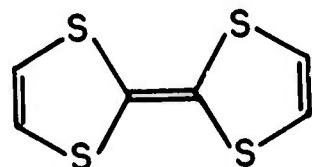
ACCEPTORS



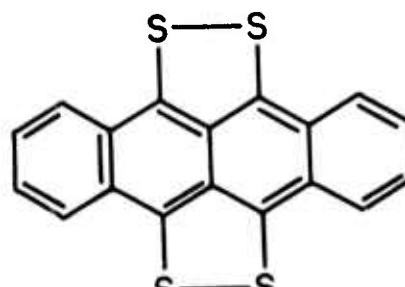
TCNQ	X = Y = H
TCNQCl	X = Cl; Y = H
TCNQBr	X = Br; Y = H
TCNQCl ₂	X = Y = Cl
TCNQBr ₂	X = Y = Br
TCNQE _t ₂	X = Y = C ₂ H ₅
TCNQCICH ₃	X = Cl; Y = CH ₃
TCNQBrCH ₃	X = Br; Y = CH ₃



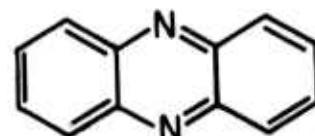
DONORS



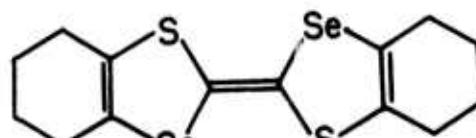
TTF



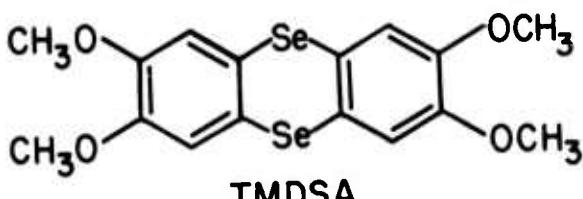
TTN



CH₃
NMP



S—Se



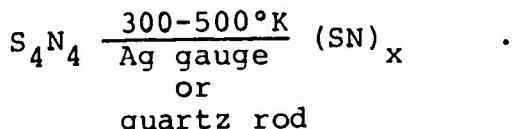
TMDSA

of complete $\sigma(\omega, T)$ measurements.

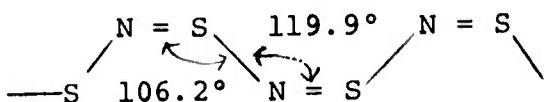
As already pointed out, the TTF mixed valence compounds like TTF-Br_{0.74} also show quite high conductivity $\sigma(\omega=0, T=300^\circ K)$ $\approx 300 (\Omega cm)^{-1}$. This decreases, with a turnover at 50°K to about $1 (\Omega cm)^{-1}$ at temperatures below 80°K. The mixed valence of 0.74 Br for each TTF gives a more stable Madelung energy through the 1:1 ratio compounds. It would be interesting to synthesize mixed valence TTF-halide (or other acceptor) systems where the acceptor is small and has a large electron affinity.

D. Polymeric Conductors

Polymeric (SN)_x was first synthesized in 1910 by F. P. Burt² utilizing the reaction



During the past two years this material has received a great deal of attention at a number of laboratories because of its remarkable physical properties, most notably that it is a highly anisotropic metal which becomes superconducting at 0.3°K. The structure of the crystal polymer as obtained by electron⁶¹ and X-ray⁶² diffraction indicates 4SN molecules per unit cell. The polymeric fibers run parallel to the b direction of the monoclinic unit cell (the b axis runs approximately along the S-N-S-N chain) with a structure given by



Since the electronegativity of N is greater than that of S, the S-N bond may have appreciable ionic character and ionic binding may therefore be significant. The structure of $(SN)_x$ is closely related to elemental tellurium, both in the average number of electrons per unit volume and in structure (Figure 18). There is still some uncertainty about the SN bond distances: the electron diffraction data suggest unequal bonds of 1.55\AA and 1.73\AA whereas the X-ray diffraction data indicate bond distances of 1.59\AA and 1.63\AA with the angles as shown above.

Measurements of the conductivity of $(SN)_x$ from $T = 4.2^\circ\text{K}$ to 300°K show highly anisotropic metallic behavior at all temperatures with a $T = 300^\circ\text{K}$ conductivity along the chain axis of $\sigma \approx 2 \times 10^3 (\Omega\text{cm})^{-1}$. The conductivity increases with decreasing temperature and exhibits a small conductivity maximum at 33°K .^{63, 65} At this meeting both the Pennsylvania group and the IBM group reported the measurement of the reflectance of $(SN)_x$ films from 500 cm^{-1} to $30,000 \text{ cm}^{-1}$. The reflectance and conductivity are highly anisotropic, but not as much as TTF-TCNQ. Metallic behavior is observed for polarization parallel to the principal conducting axis with poor metallic behavior in the perpendicular direction. The reflectance for polycrystalline $(SN)_x$ is about 45% at low frequencies and shows a well defined reflectance minimum at $22,000 \text{ cm}^{-1}$. For crystalline $(SN)_x$ the longitudinal reflectance (polarization vector parallel to the chains) may be as high as 90% below 1.5 eV.⁶⁶

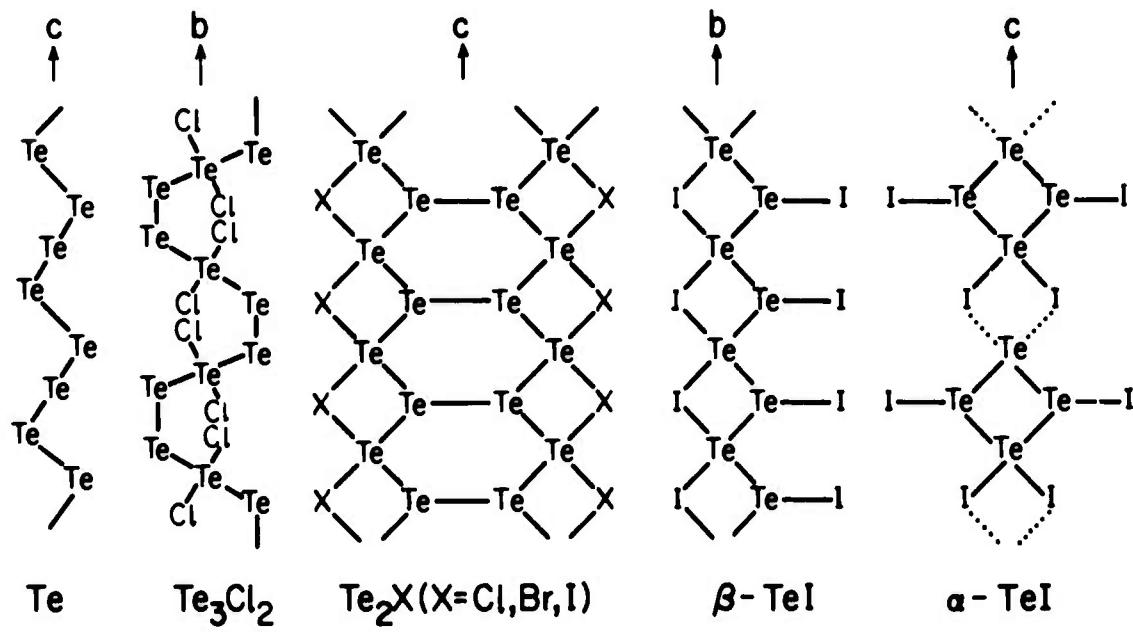


Figure 18. Structures of Te and Related Systems Derived from Te.

The lack of any sharp structure at low frequencies, aside from that at 995 and 685 cm^{-1} which may be related to S-N stretching modes, and the overall structural similarity to the reflectance of metals such as Ag^{6,7} suggests that the observed structure at 22,000 cm^{-1} is associated with a plasma edge and not with any simple electronic transition.

The reflectance data may be analyzed in a manner similar to Ag by separating the "bound" (or core) and "free" electron contributions to the complex dielectric constant $\epsilon(\omega) = \epsilon_1(\omega) + i\epsilon_2(\omega)$ into a core and Drude term:

$$\epsilon(\omega) = \epsilon_{\text{core}} - \omega_p^2 / \omega(\omega + i/\tau) .$$

Here $\omega_p = (4\pi Ne^2/m^*)^{1/2}$ is the plasma frequency, τ is the electron relaxation time, N is the electron density, and m^* is the effective mass associated with the conduction electrons. If τ can be assumed to be frequency independent, τ is the conductivity relaxation time and

$$\sigma = \omega_p^2 \tau / 4\pi .$$

By inserting the expression for $\epsilon(\omega)$ into the reflectance

$$R = \left| \frac{\sqrt{\epsilon(\omega)} - 1}{\sqrt{\epsilon(\omega)} + 1} \right|^2$$

and fitting the experimental data, Greene as well as Garito and coworkers find the constants ϵ_{core} , $\hbar\omega_p$ and τ in substantial agreement: $\hbar\omega_p \approx 5 \text{ eV}$, $\tau \approx 10^{-15} \text{ sec}$, and $\sigma \approx 4000 (\Omega\text{cm})^{-1}$. The

degree to which the values of σ obtained in this way agree with the DC value is typical of such analyses since τ is frequency dependent. Garito and Heeger⁶⁸ cite values of $N = 3 \times 10^{22} \text{ cm}^3$ and $m^* = 5.2m_e$, where m_e is the free electron mass. However, measurement of the Pauli susceptibility and the electronic specific heat⁶⁵ suggest that the density of states at the Fermi level is considerably lower than that expected from a simple band. The upturn of the reflectance above the minimum may be associated with an intra or intermolecular electronic excitation as in the case of the reflectance minimum of Ag, where interband transitions set in. This would indicate that ϵ_{core} is frequency dependent and that the plasma resonance is shifted from its conduction electron value: ω_p must therefore be viewed as a hybrid conduction-core electron resonance. Electron energy loss experiments yield a plasmon peak at the expected energy and of reasonable width.⁶⁹ In addition, there is a "superplasmon" peak at $\sim 22 \text{ eV}$ corresponding to a collective excitation of all valence electrons. The plasmon dispersion relation has also been determined experimentally. It is interesting that Landau damping is not observed in this system.

Greene reported the results⁷⁰ of OPW band calculations for $(\text{SN})_x$, which show that the Fermi energy is located at a minimum in the density of states in agreement with the susceptibility and specific heat measurements cited above.

Greene also discussed the superconductivity of $(\text{SN})_x$. $(\text{SN})_x$ has been found to be a type II superconductor with a

transition temperature $T_c \approx 0.3^\circ\text{K}$.⁵⁸ The parameter $v = \lambda/\sqrt{2}\xi$, where λ is the penetration depth, and ξ is the coherence length, is about 23. The superconductivity critical field is highly anisotropic and close to the paramagnetic limit (the field required to flip a spin). In contrast to the experience with ordinary superconductors T_c rises with increasing pressure. This is presumably associated with a stiffening of the average phonon frequency and a concomitant decrease of the electron-phonon coupling constant.

Other interesting properties of $(\text{SN})_x$ are its negative thermal expansion coefficient along the chain direction and deviations from the T^3 dependence for the lattice specific heat which is characteristic of anisotropic systems.

A point of practical interest is Garito's comment concerning the Pennsylvania group's⁵⁹ successful epitaxial growth by sublimation of $(\text{SN})_x$ on mylar, teflon, polyethylene, and others. Interesting applications are likely to emerge from this development.

III. TWO-DIMENSIONAL AND LAYERED SYSTEMS

F. DiSalvo reported on the electrical properties of layered transition metal dichalcogenides. The transition metals from the V_b group, V, Nb, and Ta(M) with S, Se, and Te(X) form interesting octahedral or trigonal prismatic type compounds which normally crystallize in the hexagonal or trigonal unit cell with a metal-metal distance near 3.3\AA .⁷¹ Each layer of MX_2 is

three atoms thick, the top and bottom layers are X atoms and M is in the middle. The atomic sheets are all regularly close-packed, so the usual notation for hexagonally packed atoms may be used. An individual layer may have XM₂ or XM₂' stacking where in XM₂ the M atoms are at the center of a trigonal prism of X and in XM₂' the M atoms are at the center of an octahedron of X atoms. The crystal structure is obtained by stacking these octahedral and trigonal layers on top of each other. The following designations indicate the stacking

1T	XM ₂ '
2H	XM ₂ , X'M ₂ '
3R	XM ₂ , X'M ₂ ', X"MX"
4Hb	XM ₂ , X'M ₂ , X'M ₂ ', X"MX"

The two dimensional 6^o layers of TX₂ are held together by Van der Waals forces.⁷²⁻⁷⁴ The layers may be chemically opened by insertion of electron donor systems; Lewis base molecules or most metals of the periodic table. This produces an intercalation compound in which a regular sequence of MX_a and donor sheets results. These layered compounds show metallic conductivity within the two-dimensional plane. The anisotropies in the resistivity can be as low as 10 or as large as 10⁵. These materials are of great physical interest since they are the first to exhibit charge density waves (CDW) of the sort first envisaged by Overhauser.⁷⁵ In addition, some of them become superconducting at low temperatures. A CDW results if

the electron-phonon interaction dominates and the Fermi surface is such that "nesting" of various sheets is possible. The two-dimensionality of these materials and their Fermi surfaces is responsible for this effect. On the other hand, if exchange forces dominate⁷⁶, a spin density wave (SDW) would result like in the antiferromagnetic state of Cr. In the following table T_0 represents the transition temperature where a CDW leads to a Peierls type distortion to an incommensurate super lattice. At a lower temperature labeled by T_d a commensurate super lattice is formed.

<u>Compound lattice (crystal type)</u>		<u>T_0 (K)</u>	<u>T_d (K)</u>
TaSe ₂	1T	600	473
TaSe ₂	2H	122	90
NbSe ₂	2H	32	none
NbS ₂	2H	no CDW	-
VSe ₂	1T	108	?

Electron and neutron diffraction^{77, 78} results clearly indicate the onset of the commensurate and incommensurate superlattice structures. The effects of impurities^{79, 80} on superconductivity in layered compounds and the addition of hydrogen atoms to layered compounds has also been studied.⁸¹

The variety of physical phenomena that are observed in these compounds is presently larger than in the one-dimensional systems, in part because the entire Fermi surface need not be destroyed by the CDW in two dimensional systems. The present

understanding of the CDW state in the two dimensional systems is more advanced than in the one dimensional systems primarily due to extensive electron and neutron diffraction studies.

IV. APPLICATIONS

Technological applications of the highly anisotropic properties of one- and two-dimensional conductors were mentioned by L. Roth, B. Scott, F. DiSalvo and others.

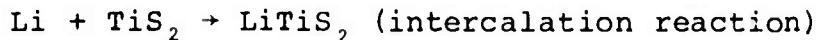
There are several applications of the highly anisotropic $(\text{SN})_x$ due to its unique ability for epitaxial growth, forming very strong bonds to well-known plastic materials. These applications include a solid state videcon, electrolytic capacitors, pyroelectric detector operation which could include polarization modulation, optical waveguide mode analysis, and by combining several of these techniques to a polarization modulated ellipsometer.

Application for the two-dimensional layered systems include the catalytic desulfuration of petroleum products, and of course the possibility of higher temperature two-dimensional superconductors. These systems, consequently, can be used as a guide to understanding CDW phenomena in one dimensional conductors. While we have emphasized the CDW instabilities of these compounds, other studies in layered compounds have been reported or are underway. These include

- a. superconductivity, particularly anisotropy of critical magnetic field H_{C_2}' ,

b. use as cathodes in non-aqueous secondary batteries.

The anode is an alkali metal, the overall reaction being for example,



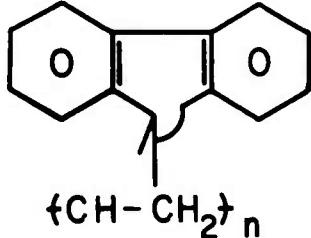
$$E_0 \sim 2.5\text{v}$$

These batteries look attractive as a possible replacement for the lead acid system,

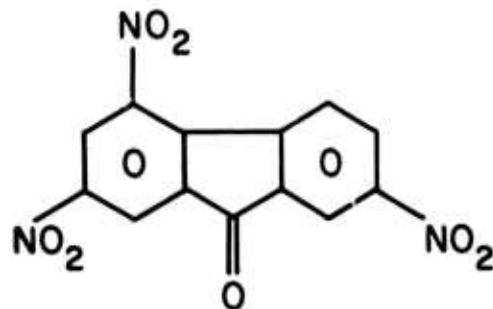
c. catalysis - $\text{Ni}_{x\text{M}}\text{S}_2$ is already used by the oil industry as a desulfurization catalyst. Studies of the newer layered compounds (substitutional alloys and intercalation compounds) should be undertaken to determine catalytic activity. Such studies have begun in several institutions, and finally

d. electrical devices - superconducting weak link detectors and other applications.

There are a large number of polymeric materials which, although not metallic, have important electronic properties such as the photoconducting polymer PVK which is used by IBM for photocopying. This polymer is used as a charge transfer complex with the acceptor TNF.



PVK (donor)



TNF (acceptor)

These systems were not discussed at the conference in detail but can be expected to provide the basis for an extensive technology in the future.

V. CONCLUSIONS AND RECOMMENDATIONS

A. Conclusions

1. Since it was last reviewed at an MRC meeting in 1973, the field of organic charge transfer and other one-dimensional conductors has made large strides. Materials preparation, the variety of experiments, and the basic physical understanding all have progressed substantially and we now have a fairly good picture of the underlying physics in one-dimensional systems.

2. It now seems unlikely that a high temperature ($T > 25K$) quasi-one-dimensional conducting material will be developed. However, exploration of this frontier of materials science may well lead to new solids with novel and presently unsuspected properties.

3. The practical applications of organic metals at present exploit only the very large conductivity anisotropies (e.g., infrared polarizers made from $(SN)_x$). The fact that broader technological uses are not yet evident should not be viewed negatively at a time when an entire new class of materials is in the beginning stages of development.

B. Recommendations

1. One-dimensional systems

a. The measurement of the complete frequency de-

pendence (DC-100GHz) of σ in TTF-TCNQ at various temperatures needs further attention in order to help decide among competing theoretical models.

b. Standard synthetic, purification, and characterization techniques should be established in order to achieve agreement on measurements in different laboratories. Exchange of samples among laboratories should be further encouraged.

c. In developing new one-dimensional systems, attention should now be given to the technological application and the relationship between structural and mechanical properties.

2. One- and Two-dimensional Systems

a. Materials should be prepared, permitting an examination of trends as we proceed from one- to two- to three-dimensional properties and the instabilities in intermediate regimes.

b. Attention should be given to synthesis of organic charge transfer systems with transversely coupled stacks. The conductivity in such systems would be less anisotropic and more nearly three-dimensional. Should such materials be discovered, the incentives for their development as opposed to the use of standard metals should be considered carefully.

3. The marriage of organic (and inorganic) synthesis and purification techniques with solid state physical methods has led to the development of new materials of high purity and properties. This approach should be exploited further. Specifically:

a. The study of organic photoconductors should be pursued. Although this topic was only mentioned briefly at this meeting, the technological applications are and undoubtedly will remain useful.

b. Of the current one-dimensional anisotropic conductors, $(SN)_x$ has the most desirable features including epitaxial growth. However, the $(SN)_x$ polymer spontaneously decomposes over a period of weeks in air. Research on new isomorphic variation on the $(SN)_x$ system should be pursued including (SeN) , $(H-C\cdot)$, and others.

||
S

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

APPENDIX

Materials Research Council (MRC) Conference on One- and Two-Dimensional Conductors Chairman, W. H. Flygare

Thursday, July 10, 1975

8:00 - 8:40	<u>G. Stucky</u> , U. of Ill. "Some Structural Aspects of One Dimensional Conductors"
9:00 - 9:40	<u>A. Heeger</u> , U. of Pa. "TTF-TCNQ and Related One-Dimensional Metals: A Status Report"
10:30 - 11:20	<u>F. DiSalvo</u> , Bell Labs., Murray Hill "Electronic Instabilities in Layered Compounds"
<u>Lunch</u>	
1:00 - 1:40	<u>R. Greene</u> , IBM, San Jose "Polymeric Conductors"
2:00 - 2:40	<u>T. Garito</u> , U. of Pa. "Studies of the Metallic Polymer (SN) _x "

Friday, July 11, 1975

8:00 - 8:40	<u>A. Bloch</u> , John Hopkins "The Design and Study of New Organic Conductors"
9:00 - 9:40	<u>B. Scott</u> , IBM, Yorktown "Organic Electronic Materials"
10:30 - 11:10	<u>L. Roth</u> , Hughes "Polymeric and Organic Conductor Applications"
<u>Lunch</u>	
1:00	Additional questions, comments and discussion

Additional Attendees: P. Anderson, Bell Labs., Murray Hill
C. Elbaum, Brown University
R. H. Baughman, Allied Chemical

MRC Members Attending this Meeting

B. Gomer, U. of Chicago	E. Montroll, U. of Rochester
F. Hawthorne, UCLA	C. Pittman, U. of Ala.
A. Bienenstock, Stanford U.	E. Ehrenreich, Harvard U.
R. Huggins, Stanford U.	J. Gilman, Allied Chemical
M. Tinkham, Harvard	M. Sinnott, U. of Mich.
E. Huckle, U. of Mich.	C. M. Stickley, ARPA
S. Ruby, ARPA	E. van Reuth, ARPA
R. Reynolds, ARPA	

REFERENCES

1. W. Knop, Ann. der phys. 43, 111 (1842).
2. F. B. Burt, J. Chem. Soc. 1171 (1910).
3. L. R. Melby, R. J. Harden, W. R. Hertler, W. Mahler, R. E. Benson and W. E. Mochel, J. Am. Chem. Soc. 84, 3374 (1962).
4. C. J. Fritche, Acta Cryst. 20, 892 (1966).
5. N. F. Mott, Proc. Phys. Soc. (London) A62, 416 (1949).
6. E. H. Lieb and F. Y. Wu, Phys. Rev. Lett. 20, 1445 (1968).
7. R. E. Peierls, "Quantum Theory of Solids." (Oxford University Press, London, 1955).
8. P. A. Lee, T. M. Rice and P. W. Anderson, Solid State Commun. 14, 703 (1974).
9. H. Mook as reported by A. J. Heeger, MRC Conference (1975).
10. R. A. Gardner, N. Blasse, and A. Wold, Acta Crist. B25, 281 (1969).
11. J. B. Goodenough, Sol. State Chem. 5, 145 (1971) and 12, 376 (1975).
12. L. V. Interrante, K. W. Browall, and S. P. Bundy, Inorg. Chem. 13, 1158 (1974) and L. Interrante and K. W. Browall, Inorg. Chem. 13, 1162 (1974).
13. H. Zeller, Festkörperprobleme 13, 31 (1973).
14. K. Krogman, Angew. Chem. Int. Ed. English 8, 35 (1969).
15. D. Kuse and H. R. Zeller, Phys. Rev. Lett. 27, 1060 (1971).
16. D. Kuse and H. R. Zeller, Solid State Comm. 11, 355 (1972).
17. P. Stampli, Phys. State. Sol. A10, 537 (1972).
18. F. N. Lecrone and J. H. Perlstein, Chem. Comm. 75, (1972).
19. M. J. Minot and J. H. Perlstein, Phys. Rev. Lett. 26, 371 (1971).

20. J. M. Williams, J. L. Peterson, H. M. Gerdes, and S. W. Peterson, Phys. Rev. Lett. 33, 1079 (1974).
J. M. Williams, M. Iwata, F. K. Ross, J. L. Peterson, and S. W. Peterson, Mol. Res. Bull. 10, 411 (1975).
21. R. Comes, M. Lambert, H. Launois and H. R. Zeller, Phys. Rev. B8, 571 (1973).
22. B. Renker, L. Pintechouine, W. Gläser, H. Rietschel, R. Comes, L. Liebert and W. Drexel, Phys. Rev. Lett. 32, 836 (1974).
23. B. Renker, H. Rietschel, L. Pintschovies, W. Gläser, P. Griesch, D. Kuse and M. J. Rice, Phys. Rev. Lett. 30, 1144 (1973).
24. J. W. Lynn, M. Iizumi, G. Shiane, S. A. Werner and R. B. Saillant, to be published.
25. K. Krogman and H. P. Geserich, ACS Symposium Series 5, 350 (1974).
26. A. F. Garito and A. J. Heeger, Accts. Chem. Res. 7, 232 (1974); see also earlier references.
27. A. J. Heeger and A. F. Garito, A.I.P. Conf. Proc. 10, 1476 (1973).
28. A. Bloch, D. Cowan and T. O. Poehler, "Energy and Charge Transfer in Organic Semiconductors," Plenum Press, N. Y. (1974).
29. T. E. Phillips, T. J. Kistenmacher, J. P. Ferraris and D. O. Cowan, J.C.S. Chem. Comm. 471 (1973) and Acta. Cryst. B30, 763 (1974). See also R. H. Blessing and P. Coppens, Solid State Commun. 15, 215 (1974).
30. J. P. Ferraris, D. O. Cowan, V. Walatka, Jr. and J. H. Perlstein, J. Amer. Chem. Soc. 95, 948 (1973); J. H. Perlstein, J. P. Ferraris, V. V. Walatka, Jr. and D. O. Cowan, Proc. 18th Conference on Magnetism and Magnetic Materials, Ed. by C. D. Grahem and J. J. Rhyne; A.I.P. Conference Proc. 10, 1494 (1972).
31. L. B. Coleman, M. J. Cohen, D. J. Sandman, F. G. Yamagishi, A. F. Garito and A. J. Heeger, Solid State Commun. 12, 1125 (1973).
32. D. E. Schafer, F. Wudl, G. A. Thomas, J. P. Ferraris, and D. O. Cowan, Solid State Commun. 14, 342 (1974).

33. M. J. Cohen, L. B. Coleman, A. F. Garito and A. J. Heeger, Phys. Rev. B10, 1298 (1974); T. Wei, S. Etemad, A. F. Garito and A. J. Heeger, Phys. Lett. 45a, 269 (1973).
34. G. A. Thomas, et. al., Solid State Comm. 00, 000 (1975). See also A. J. Heeger and A. F. Garito, preprint entitled "Comment on the Electrical Conductivity of TTF-TCNQ."
35. R. S. Swingle, R. P. Groff and B. M. Monroe, Phys. Rev. Lett. in press (1975).
36. A. N. Bloch, J. P. Ferraris, D. O. Cowan, and T. O. Poehler, Sol. State Comm. 13, 753 (1973).
37. S. K. Khanna, E. Ehrenfreund, A. F. Garito and A. J. Heeger, Phys. Rev. Lett. B10, 2205 (1974).
38. D. B. Tanner, C. S. Jacobson, A. F. Garito and A. J. Heeger, Phys. Rev. Lett. 32, 1301 (1974).
39. A. A. Bright, A. F. Garito and A. J. Heeger, Solid State Comm. 13, 948 (1973).
40. R. A. Craven, M. B. Salamon, G. DePasquali, R. M. Herman, G. Stucky and A. Schultz, Phys. Rev. Lett. 32, 601 (1973).
41. P. M. Chaikin, J. F. Kwak, T. E. Jones, A. F. Garito and A. J. Heeger, Phys. Rev. Lett. 31, 601 (1973).
42. P. I. Perov and J. E. Fisher, Phys. Rev. Lett. 33, 521 (1974).
43. A. Bloch, reported at La Jolla meeting.
44. M. B. Salamon, J. W. Bray, G. DePasquali, R. A. Gaven, G. Stucky and A. Schultz, Phys. Rev. B11, 619 (1975).
45. M. J. Cohen, L. B. Coleman, A. F. Garito and A. J. Heeger, Phys. Rev. B10, 1298 (1974).
46. S. K. Khanna, E. Ehrenfreund, A. F. Garito and A. J. Heeger, Phys. Rev. B10, 2205 (1974).
47. J. C. Scott, A. F. Garito and A. J. Heeger, Phys. Rev. B10, 3131 (1974).
48. F. Denoyer, R. Comes, A. F. Garito and A. J. Heeger, Phys. Rev. Lett. 35, 445 (1975).
49. A. J. Heeger and A. F. Garito, Low Dimensional Cooperative Phenomena, Ed. by H. J. Keller (Plenum Press, N.Y. and London, 1975), p. 89.

50. P. A. Lee, T. M. Rice and P. W. Anderson, Phys. Rev. Lett. 31, 462 (1973).
51. S. K. Khanna, A. F. Garito, A. J. Heeger and R. C. Jaklevic, Sol. State Comm. 16, 667 (1975).
52. R. G. Merrifield, Phys. Rev. Lett. 34, 877 (1975).
53. V. Bernstein, P. M. Chaikin and P. Pincus, Phys. Rev. Lett. 34, 271 (1975).
54. S. Etemad, T. Penny, E. M. Engler, B. A. Scott and P. E. Seiden, Phys. Rev. Lett. 34, 741 (1975).
- 54a. A. Luther and I. Peschel, Phys. Rev. Lett. 32, 992 (1974).
- 54b. J. Black and A. Luther (to be published).
55. A. J. Berlinsky, J. F. Carolan and L. Weiler, Solid State Comm. 15, 795 (1974).
56. A. N. Bloch, D. O. Cowan and T. O. Poehler, Tech. Report to ARPA, July 7, 1975.
57. T. O. Poehler, A. N. Bloch, J. P. Ferraris and D. O. Cowan, Solid State Comm. 5, 337 (1974).
58. Y. Tomkiewicz, B. A. Scott, L. J. Tao and R. S. Title, Phys. Rev. Lett. 32, 1363 (1974).
59. A. N. Bloch, J. P. Ferraris, D. O. Cowan and T. O. Poehler, Solid State Comm. 13, 753 (1973).
60. A. N. Bloch, D. O. Cowan, K. Bechgaard, R. E. Pyle, R. H. Banks and T. O. Poehler, preprint.
61. M. Boudeulle and P. Michel, Acta Cryst. Sect. A28, S199 (1972).
62. M. J. Cohen, A. F. Garito, A. J. Heeger, M. S. Saran, C. M. Mikulski and G. G. MacDiarmid, J. Am. Chem. Soc. 00, 0000 (1975).
63. V. V. Walatka, Jr., M. M. Labes and J. H. Perlstein, Phys. Rev. Lett. 31, 1139 (1973); C. H. Hsu and M. M. Labes, J. Chem. Phys. 61, 4640 (1974).
64. A. A. Bright, M. J. Cohen, A. F. Garito, A. J. Heeger, C. M. Mikulski, P. J. Russo and A. G. MacDiarmid, Phys. Rev. Lett. 34 206 (1975).

65. R. L. Greene, P. M. Grant and G. B. Street, Phys. Rev. Lett. 34, 89 (1975); R. L. Greene, G. B. Street and L. J. Suter, Phys. Rev. Lett. (to be published). Apparently the better quality crystals do not show the maximum at 33°K.
66. A. A. Bright, M. J. Cohen, A. F. Garito, A. J. Heeger, C. M. Mikulski and A. G. MacDiarmid, to be published.
67. H. Ehrenreich and H. R. Philipp, Phys. Rev. 128, 1622 (1962).
68. A. J. Heeger and A. F. Garito, Semi-Annual Technical Report to ARPA, January 1, 1975.
69. C. H. Chen, J. Silcox, A. F. Garito, A. J. Heeger and A. G. MacDiarmid, Phys. Rev. Lett. (submitted).
70. H. Kaminuro, A. J. Grant, F. Levy, A. D. Yoffe and G. D. Pitt, Solid State Comm. 17, 49 (1975) and D. E. Parry and J. M. Thomas, J. Phys. C8, 645 (1975). W. E. Rudge and P. M. Grant, Phys. Rev. Lett. 00, 0000 (1975).
71. F. J. DiSalvo, B. G. Bagley, J. M. Voorhoeve and J. V. Waszczak, J. Phys. Chem. Sol. 34, 1357 (1973).
72. F. J. DiSalvo, G. W. Hull, L. H. Schwartz, J. M. Boorhoeve and J. V. Waszczak, J. Chem. Phys. 59, 1922 (1973).
73. F. R. Gamble, J. H. Osiecki and F. J. DiSalvo, J. Chem. Phys. 55, 3525 (1971).
74. F. R. Bamble, J. H. Osiecki, M. Cais, R. Pishardy, F. G. DiSalvo and T. H. Geballe, Science 174, 493 (1971).
75. A. W. Overhauser, Phys. Rev. 167, 691 (1968) and B3, 3173 (1971).
76. J. A. Wilson, F. J. DiSalvo and S. Mahajon, Phys. Rev. Lett. 32, 882 (1974). P. M. Williams, G. S. Parry and C. B. Scruby, Phil. Mag. 29, 695 (1974).
77. D. B. Moncton, J. D. Axe and F. J. DiSalvo, Phys. Rev. Lett. 34, 734 (1975).
78. J. A. Wilson, F. J. DiSalvo and S. Mahajon, Adv. in Phys. 24, 117 (1975).
79. J. J. Hauser, M. Robbins and F. J. DiSalvo, Phys. Rev. B8, 1038 (1973).

80. F. J. DiSalvo, J. A. Wilson, B. G. Bagley and J. V. Waszczak, Phys. Rev. B, 0000 (1975).
81. D. W. Murphy, F. J. DiSalvo, G. W. Hull, J. V. Waszczak, S. F. Mayer, G. R. Stewart, S. Early, J. V. Acrivos and T. H. Geballe, J. Chem. Phys. 62 967 (1975).

ANTENNA STRUCTURES FOR INFRARED DIODES

P. L. Richards and A. Yariv

ABSTRACT

The problem of efficiently coupling the energy from a free space infrared beam to a diode with dimensions less than one wavelength has been studied. Contributions under the subject of optimizing the feed structures for high gain antennas include:

1. Estimates of the efficiency of presently used schemes.
2. Suggestions of improved single antenna structures.
3. Analysis of array feed structures.

Contributions under the subject of impedance matching include:

1. Fundamental limits to impedance matching.
2. Effect of antenna resistance, transmission line resistance, and diode series resistance on impedance matching and speed of response.
3. Use of arrays for impedance matching.

It appears that the techniques now generally used to couple to MOM diodes can easily be improved by about one order of magnitude even at low infrared frequencies. A second order of magnitude might be available with a carefully optimized array.

ANTENNA STRUCTURES FOR INFRARED DIODES

P. I. Richards and A. Yariv

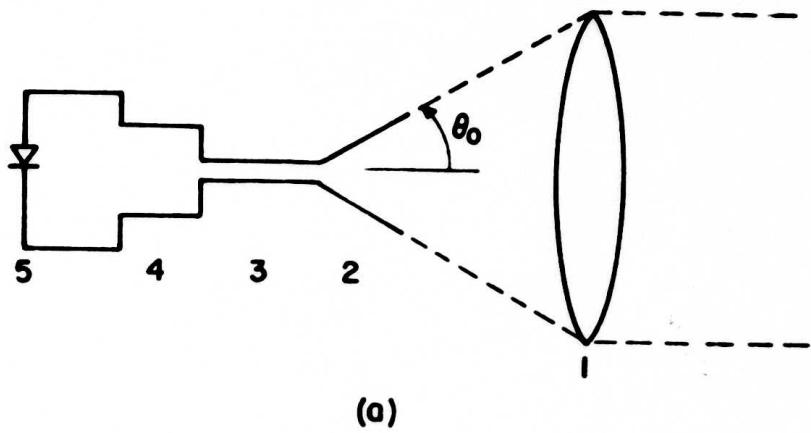
There has been great interest in extending the range of lumped circuit nonlinear elements to infrared and optical frequencies. The highest frequency work has used metal-oxide-metal (MOM) diodes at room temperature,¹⁻⁵ but there is also considerable interest in high frequency applications of Josephson junctions⁶ as well as both conventional and superconducting Schottky barrier diodes.^{7,8} Because transmission lines have relatively large losses at most infrared frequencies it will often be necessary to couple electromagnetic radiation propagating in free space to the diode structure.

In applications such as low noise receivers, it is vital to achieve efficient coupling. Coupling efficiency is also important for a variety of other frequency conversion experiments because large amounts of power are available only at a limited number of infrared frequencies. Thus far, however little attention has been paid to the problem of coupling efficiency in the infrared frequency range.

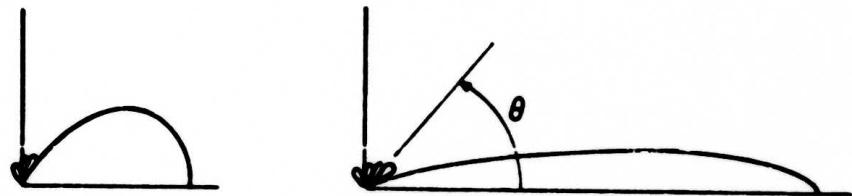
In principle it is possible to couple all of the power in any single mode of electromagnetic radiation into a diode of any impedance. In practice there are limitations at any frequency, and many of these are especially severe at infrared

and optical frequencies. It is traditional to divide the coupling problem into an antenna pattern problem and an impedance matching problem, although the general concept of impedance could be extended sufficiently to cover the whole. A typical system for coupling a collimated beam to a diode is illustrated in Fig. 1. It consists of a primary optical element or collimator which can be a lens or a paraboloidal reflector, a feed structure to collect the focused power, and a transmission line which is matched to both the feed and to the diode. In this paper we will primarily consider the antenna problem. Some aspects of the impedance problem will be considered in a separate report.

There is a well developed technology for coupling a free space microwave frequency beam to a transmission line.⁹ Many of the concepts and some of the techniques developed for that purpose can be carried over directly to the infrared. The well known reciprocity theorems⁹ relate the properties of receiving and transmitting antennas, so that we can speak of them interchangeably. The far-field intensity pattern of an antenna is characterized by a gain function $G(\theta, \phi)$ which is normalized such that $\int_{4\pi} G(\theta, \phi) d\Omega = 4\pi$. This gain function is related to the antenna differential cross section $S(\theta, \phi) = G(\theta, \phi) \lambda^2 / 4\pi$. The physical idea behind this relation can be expressed by the statement that for a single spatial mode, the product of the effective area A of a beam times its solid angle of divergence Ω is an invariant quantity equal to λ^2 .



(a)



(b)

Figure 1(a). Schematic diagram of diode receiver consisting of a primary element 1 shown as a lense, a feed structure 2 matched to a transmission line 3, a matching structure 4, and a diode 5.

1(b). Schematic gain curves for the receiver (left) and the feed structure (right).

The spatial distribution of the mode of interest determines the details of the required antenna structure. Two cases of practical importance will be considered. In a radar receiver or an astronomical telescope the transverse spread of the incoming mode is so large that any practical primary element is uniformly illuminated. In a typical laboratory experiment with a single mode TEM_{00} laser, on the other hand, the illumination of the primary is Gaussian.¹⁰ If we consider the case of a transmitting antenna, it is clear that the gain function $G(\theta)$ of the optimum feed structure will be square or Gaussian in the two cases. The theorems on reciprocity and integrated cross section mentioned above show that this result is also valid for the receiving antenna.

For a typical feed structure, $G(\theta, \phi)$ will have a central lobe flanked by several side lobes. The phase of this antenna pattern (referred to a spherical wave front) changes by $\sim\pi$ from the peak of one lobe to the next. In general, therefore, a calculation of the coupling to an incoming mode will require phase information about the feed structure which is difficult to measure at infrared frequencies.

We will now describe an approximate method of computing the coupling which neglects phase cancellation. In this picture the incoming radiation is focused by the primary onto a spot of effective area A . The dimensions of the feed structure are assumed to be small enough that the intensity $I(\theta, \phi)$ is independent of position and depends only on angle. The intensity

in each element of solid angle $d\Omega$ is intercepted by the corresponding differential cross section $S(\theta, \phi)$ of the feed structure. If the radiation resistance of the feed structure is matched to a resistive load, the power delivered to that load is $I(\theta, \phi)$ times $S(\theta, \phi)$ integrated over solid angle. If we divide by the power P_0 illuminating the primary, which is $I(\theta, \phi)$ integrated over solid angle times the area A of the beam, then we have an estimate of the coupling efficiency,

$$\epsilon = \frac{\lambda^2}{4\pi} \int_{4\pi} G(\theta, \phi) I(\theta, \phi) d\Omega / A \int_{4\pi} I(\theta, \phi) d\Omega \quad (1)$$

For the uniform illumination mode focused into a solid angle Ω_0 , the effective area at the focus is $A = \lambda^2/\Omega_0$. For an intercepted power P_0 , $I(\theta, \phi) = P_0/A\Omega_0 = P_0/\lambda^2$ for angles $\theta < \theta_0$. We have implicitly assumed that the diameter of the primary D is large enough that we can neglect any variation in $G(\theta, \phi)$ over the spread of angles $\approx \lambda/D$ which arise from diffraction at the primary.

It is convenient to characterize a focused Gaussian beam in terms of the radius w_0 at the waist.¹⁰ The effective "pillbox" area is $A = \pi w_0^2/2$. If we decompose the intensity distribution at the waist into plane waves we find the exact result, $I(\theta) = (4P_0/\lambda^2) \exp(-2\pi^2\theta^2w_0^2/\lambda^2)$.

Using these values for $I(\theta)$ and A , it is possible to estimate the coupling efficiency for any feed structure whose $G(\theta, \phi)$ has been measured or calculated from theory. Since

phase cancellation has been neglected, the coupling computed from (1) is an upper limit. The approximation is least serious for a feed structure with a Gaussian $G(\theta)$ because the wave front is then spherical. The estimate is accurate for any feed structure whenever the primary only accepts radiation from well within the central lobe of $G(\theta, \phi)$. Unfortunately, the coupling efficiency is always poor under such circumstances. In cases where several lobes are important, the estimate (1) can be improved by carrying out the calculation separately over the various lobes and summing the results with alternating signs. Since the gain function $G(\theta, \phi)$ of useful feed structures is often strongly dependent on polarization, the distribution of polarizations in $I(\theta, \phi)$ must be considered in any complete calculation. Other factors which can reduce the efficiency achieved in practice are resistive dissipation in the feed and aberrations or poor optical quality of the primary. A primary reflector, for example, with mean square random deviation σ^2 from the ideal surface, will reduce ϵ by a factor $\approx \exp 2\pi\sigma^2/\lambda^2$.

In the following sections we discuss the efficiency of several coupling schemes and comment on their usefulness at infrared frequencies.

1. Near-isotropic feeds

If the feed pattern $G(\theta, \phi)$ is slowly varying over the range of angles with significant $I(\theta, \phi)$, then $G(\theta, \phi)$ can be replaced by its maximum value G (the feed gain) in Eq. (1). For the case of uniform illumination we then have $\epsilon = G\Omega_0/4\pi$.

For an isotropic feed, $G = 1$ and ϵ is just the fraction of the total solid angle subtended by the primary. For an f/l system, for example, $\epsilon = 0.08$. For the case of Gaussian illumination, the efficiency is $\epsilon = G\lambda^2/2\pi^2W_0^2$.

2. Dipole Feeds

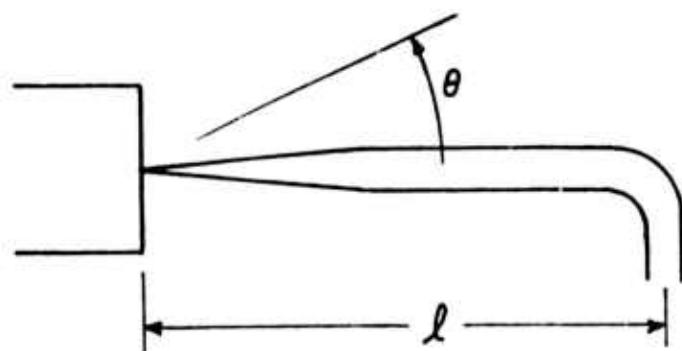
The half wave dipole is an antenna of special interest because it can be easily deposited in connection with a thin film diode.¹¹ If we neglect the substrate interaction, the gain peaks at the value $G = 1.65$ for directions perpendicular to the dipole. A ground plane deposited on the reverse side of a substrate of optical thickness $\lambda/4$ will increase the gain by a factor ~ 4 . The general features of the gain curve $G(\theta, \phi)$ of such an antenna are well known⁹ since it is used as an antenna feed at microwave frequencies. It is essentially that of a quadrupole, so a rather fast primary element is required for best coupling. If the roll-off of the gain curve is adjusted to match that of a Gaussian beam, quite appreciable coupling efficiencies appear possible. The coupling will be less good for the uniform illumination mode.

Because of the simplicity of this structure it is likely to have important applications. One can imagine, for example, an infrared imaging system consisting of a fast primary with an array of dipole antennas (backed with a ground plane) at the focus. Each dipole would have its own diode and¹¹ output circuitry.

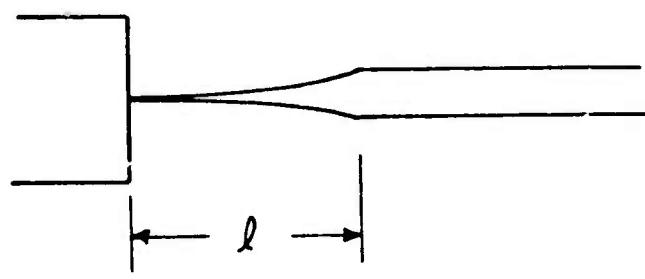
3. Long Wire Antenna Feed

The conventional point contact MOM diodes consist of a sharpened wire pressed against a flat anvil as shown in Fig. 2 (a,b). The observed antenna pattern can usually be described by applying long wire antenna theory^{1,2,13} to the straight portion of the wire. For this antenna $G(\theta, \phi)$ has the axial symmetry of the wire and rises from zero at $\theta = 0$, has a peak at $\theta_p = \cos^{-1}(1-0.37\lambda/\ell) \approx 0.86\sqrt{\lambda/\ell}$ for $\theta \ll 1$ followed by a series of subsidiary maxima as shown in Fig. 3. Deviations from long wire antenna theory occur (presumably) due to the wire diameter exceeding one wavelength. For the antenna shown in Fig. 2(b) these reduce θ_p by a factor of order 2.¹³

It is remarkable that any agreement with long wire antenna theory is seen under these conditions. One can speculate that surface wave propagation along the wire becomes important and that the effective length of the wire is determined by the launching of surface waves in regions of convex curvature. For the case of a surface which has a curvature of radius R along the direction of propagation, but is flat in the perpendicular direction, the rate at which surface mode intensity is launched is given by¹⁴ $d \ln I/d\theta = -3(R/\lambda)^{1/3}$. Here θ is the angle through which the surface is bent. For a typical antenna of type 2(a) we might have $(R/\lambda)^{1/3} \approx 1$, so that a 90 degree bend is adequate to launch the surface modes. It is interesting to speculate whether there is traveling or



(a)



(b)

Figure 2. Two types of long wire antennas with effective length l .

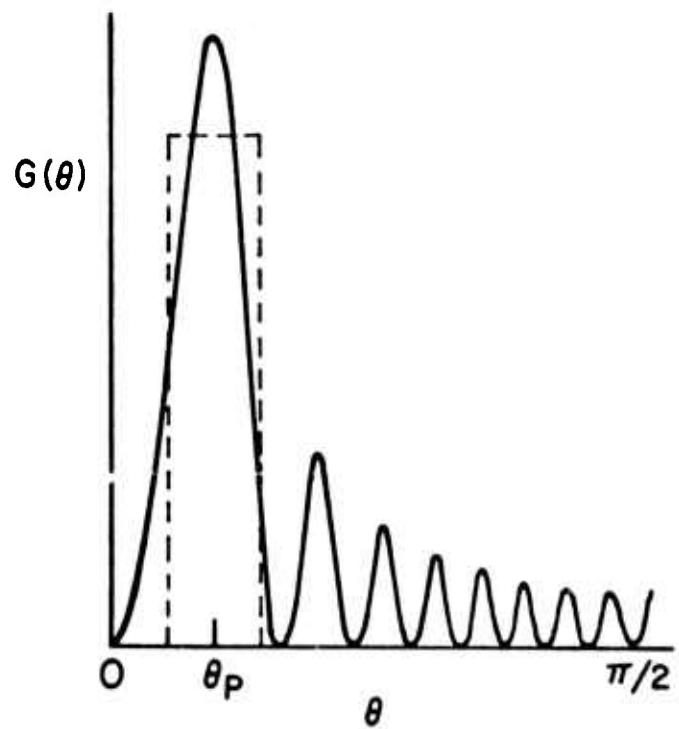


Figure 3. Box approximation to a long wire antenna gain curve.

standing wave excitation of the wire. It is difficult to understand how sufficiently strong reflections could occur at a long radius bend of the type shown in Fig. 2(a) for standing waves to dominate. This seems more reasonable in case (b) where the curvature is localized within one wavelength. An analysis of surface waves on a conical antenna has been reported which may help in understanding antennas of the type (b).¹⁵ It appears from the reported data that the anvil usually functions as a ground plane since strong lobes are seen only for small angles θ as defined in Fig. 2(a).

In order to discuss the efficiency of coupling to such an antenna without resorting to numerical calculations we replace the real curve $G(\theta)$ with a box extending from $\theta_p/2$ to $3\theta_p/2$ as shown in Fig. 3(a). The gain of this feed is easily computed from the area of a sphere divided by twice the area of the segment of the sphere which extends from $\theta_p/2$ to $3\theta_p/2$. We obtain $G = \frac{1}{2}(\cos \theta_p/2 - \cos 3\theta_p/2)^{-1} \approx 1/2\theta_p^2$, for $\theta_p \ll 1$. The usual practice is to couple this pattern with a primary lens whose axis lies along $\theta_p^{1/2}$, although it has been suggested that better performance can be obtained with a primary whose axis lies along the wire axis.^{16,17} Our approximate model supports this idea. The fraction of the box antenna pattern intercepted by a lens of diameter θ_0 whose axis lies along θ_0 is $\epsilon \approx (1-\cos \theta_p/2)/(\cos \theta_p/2 - \cos 3\theta_p/2) \approx 1/8$ when $\theta_p \ll 1$. By contrast, if the length of the wire is increased by a factor ~ 4 so that the same lens looking along the wire intercepts all

angles out to $3\theta_p/2$, then the entire box pattern at that end of the antenna is intercepted. The efficiency is then $\epsilon = 4/9$. This includes a factor $(\cos \theta_p/2 - \cos 3\theta_p/2)/(1 - \cos 3\theta_p/2)$ $\approx 8/5$ when $\theta_p \ll 1$, for the fact that there is no radiation in the cone of half angle $\theta_p/2$, and 1/2 because by symmetry half of the focused beam has the wrong polarization to interact with the wire. Polarization effects also reduce the efficiency of coupling to a primary whose axis lies along θ_p , but by a smaller amount.

Despite the crude nature of this estimate, it suggests that there may be an important improvement in performance available in the end-on geometry. The length of the wire used in practice is influenced by such factors as the shadowing by the wire itself, the instability of very long wires, and the speed of the available primary optics.

4. Feed Horns

One of the most useful feed structures at microwave frequencies is a metal waveguide termination in a horn. Such horns are well understood and efficient antenna feeds can be designed.⁹ The horns have rectangular cross sections with sides oriented with respect to the polarized input beam to minimize mode conversion. Unwanted modes are reflected by a length of fundamental mode waveguide (usually TE_{10}) matched to the diode. The problem of fabricating a matched section of single mode hollow metal infrared waveguide around the diode makes this approach useful only for the longest far

infrared wavelengths. The problem is especially severe if an intermediate frequency signal must be extracted efficiently.

If an open diode structure is used, it must be located very close to the end of a waveguide or to the small end of the horn to avoid finding itself in a diverging field. The coupling to the diode in this case can only be estimated by computing an overlap integral since near-field effects will be important.

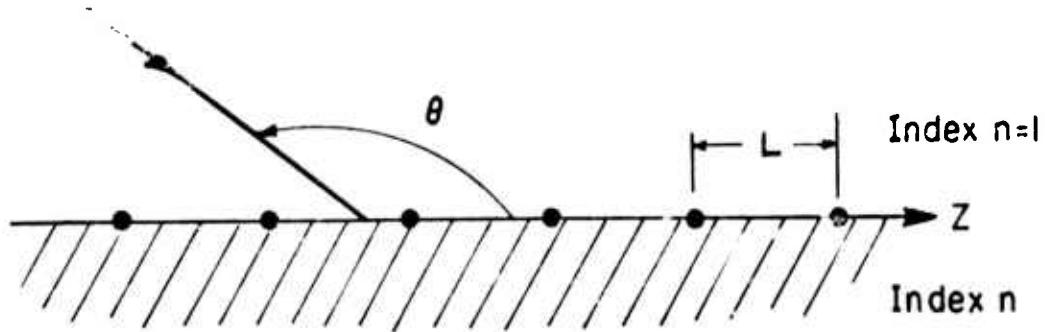
Conical light concentrators have occasionally been used to feed point contact junctions. This is a very poor procedure for use with single mode sources because of the large mode conversion expected. The conical concentrator is not an imaging device, even in the limit of geometrical optics. If a fraction f of the radiation from a single mode source is converted into n equally populated modes, and if we assume that n' of these modes arrive at the diode within the acceptable solid angle, but with random phases, the coupled power should be reduced by a factor $[1+f(\sqrt{n'}/n-1)]$. This objection does not apply to a cone which is fed from a multimode source such as a black body. If all relevant modes are fully populated then mode conversion is unimportant. No matter what its gain curve, any antenna which is completely immersed in black body radiation is coupled with unit efficiency.

5. Array Antennas

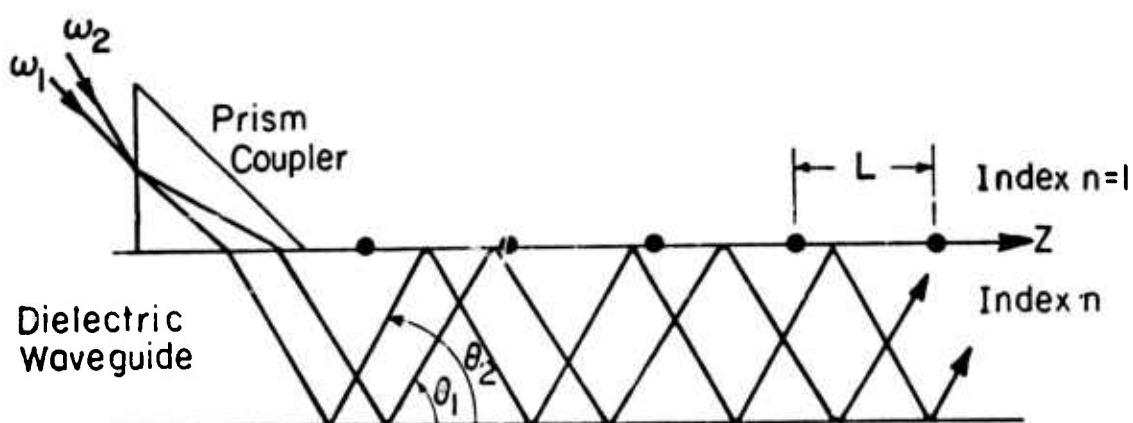
Arrays of straight wire antennas can be fabricated relatively easily in connection with deposited junctions.

The properties of such arrays have been studied for applications in lower frequency regions. At radio frequencies where parabolic reflectors are cumbersome, arrays are used to obtain very high gain. It seems probable that infrared practice will be more like that in the microwave frequency range where primary antennas are usually used to obtain high gain. For this reason we have focused our attention on useful feed structures for lens or reflecting parabolic primaries. It seems possible, however, that important applications for arrays containing infrared diodes may be found. It is of interest therefore to explore some possible ways in which they may be used.

An array of N wires lying along the x direction, and spaced a distance L apart in the z direction as is shown in Fig. 4(a) has many well known and interesting properties. The gain curve in the $x-y$ plane is roughly given by the single wire antenna pattern (e.g., dipole or long wire). The gain curve in the $y-z$ plane can often be guessed from the analogy to an N -slit diffraction grating. The gain of any of the major lobes is proportional to N^2 . If the array is fed with constant phase there will be lobes of order $m = 0, \pm 1, \pm 2$ at angles which are symmetric with respect to the $x-y$ plane. If the angle between the free space wave vector and the z axis is θ , then the diffraction condition is $(\omega/c)\cos \theta = 2\pi m/L$. The same equation governs a set of orders of diffraction that propagate into the substrate of index n , except that in this case c is replaced by c/n . To illustrate the potential use



(a)



(b)

Figure 4 (a). Coupling from a free space wave into an antenna array.

4 (b). Coupling from two free space waves into a diode array by means of a dielectric waveguide.

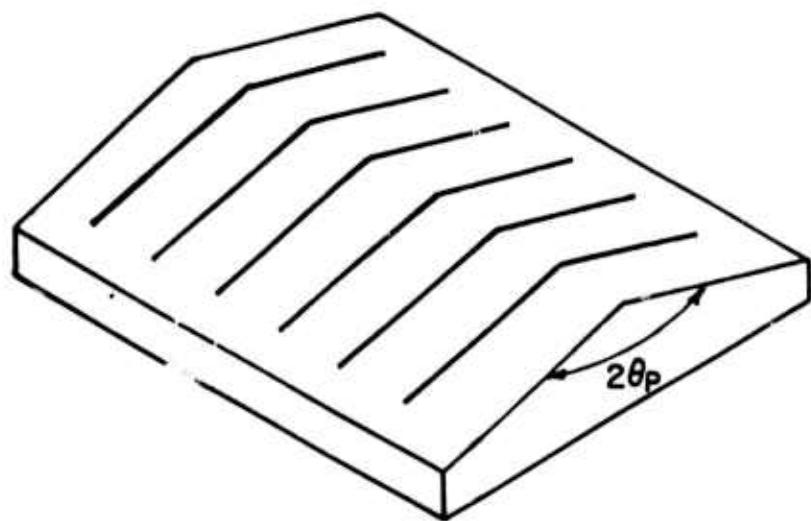
of such a structure, we will consider a frequency conversion experiment. Each wire can function as a resonant long wire antenna at 3 frequencies if $l = p\lambda_1 = q\lambda_2 = r\lambda_3$, where p, q, and r are integers. If a diode which requires no dc bias is located along each wire at a point of maximum current for all modes, a relatively efficient frequency converter will result if the available power is sufficient to drive the junctions properly. The efficiency is maximized by selecting a wire separation close enough to suppress the first order of diffraction. In the y-z plane the peaks of the gain function correspond to reflection and transmission of a homogeneous sheet (zero order of diffraction). In the y-z plane the gain function will have two peaks for p, q , and $r \gg 1$. This type of antenna, therefore, has a minimum of 4 comparable well defined lobes. Since it is possible to couple to only one of these lobes in practice, the efficiency of coupling cannot be greater than $1/4$ that of a single lobe feed structure. For a system which must both receive and transmit this becomes a very significant factor of $1/16$.

It appears likely that some of the deficiencies of this structure can be remedied by using "V antennas", rather than straight long wire antennas. The V antenna has a two lobe pattern oriented along the axis of the V provided that the angle between the legs of the V is equal to $2\theta_p$ of the individual long wire antennas.¹⁸ The coupling efficiency could therefore be increased by a factor 2 if the wires were

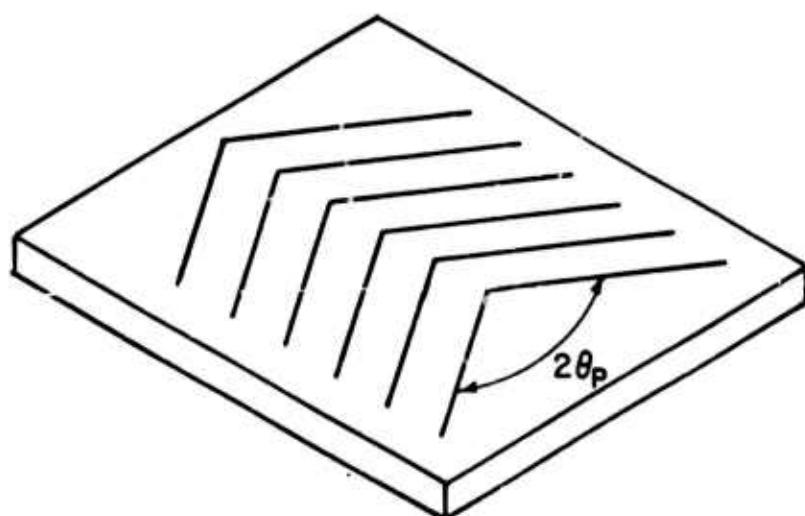
deposited on a substrate with the appropriate dihedral angle as is shown in Fig. 5(a). Alternatively the chevron pattern in Fig. 5(b) could be used near grazing incidence to make an antenna analogous to the common television antenna.

It will be difficult in general to achieve the case of a single dominant lobe with approximate axial symmetry which is most useful for efficient coupling to a free space wave. In the radio region this is done with end-fire traveling wave arrays⁹ which are fed with transmission lines with suitable phase delays, or inductively as in the Yagi antennas¹⁹. The velocity of propagation of either type of feed wave will be given by an effective index of refraction n' which is intermediate between that of the substrate and free space. Because of this substrate interaction it is not clear how to construct true end-fire arrays. It is still possible, however, to couple the array to free space waves using the standard techniques of phase matching. If the free space wave approaches at an angle θ as shown in Fig. 4(a) then conservation of wave vector in the z direction is accomplished with an array of spacing L if $(\omega/c)\cos \theta = 2\pi m/L + n'\omega/c$, where m is the order of diffraction. Since $|\omega/c)\cos \theta| < n'\omega/c$ for the free space wave, the lowest order of phase matching to the guided wave which determines the antenna phases is for $m = -1$. Zero order forward phase matching is possible for the wave which propagates into the substrate.

At certain frequencies in the infrared it will be possible to construct low loss dielectric waveguides. A prism can



(a)



(b)

Figure 5(a). Array of "V" antennas with main lobes perpendicular to the plane of the substrate.

5(b). Array of "V" antennas with main lobes in the plane of the substrate.

be used as shown in Fig. 4(b) to couple a free space wave into such a waveguide with efficiencies approaching 0.8. It may be useful to couple a free space wave to an array of diodes by means of such a waveguide.

Let us consider the problem of mixing a wave at frequency ω_1 with a wave at ω_2 in a line of diodes arrayed along a dielectric waveguide. We will assume that each diode is provided with a "coupler" which couples it weakly to the traveling modes in the guide. The leads to the diodes may well be adequate for this purpose. The output at $\omega_2 - \omega_1 \ll \omega_1$ is coupled out along a strip transmission line to which the diodes are attached. To get cumulative reinforcement of all the diode outputs we must satisfy the phase matching condition

$$(n_2 \omega_2 / c) \cos \theta_2 - (n_1 \omega_1 / c) \cos \theta_1 = n'_3 (\omega_2 - \omega_1) / c + 2\pi m / L, \quad (2)$$

where n_1 and n_2 are indices of refractions, and θ_1 and θ_2 are the internal "zig-zag" angles of the dielectric wave guide modes at ω_1 and ω_2 , and n' is the effective index of refraction at ω_3 .

Note that for no dispersion ($n_1 = n_2 = n_3$) Eq. (2) would be satisfied with $\theta_1 = \theta_2 = 0$. It follows that values of $L \gg \lambda$ will be used to phase match. Since the array of coupled diodes will in principle act as a high order grating coupler which will radiate to free space waves, the diode couplers should be designed to minimize such effects.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

References:

1. L. O. Hocker, D. R. Sokoloff, V. Daneu, A. Szoke, and A. Javan, *Appl. Phys. Lett.* 12, 401 (1968).
2. V. Daneu, D. R. Sokoloff, A Sanchez and A. Javan, *Appl. Phys. Lett.* 15, 398 (1969).
3. C. Freed and A. Javan, *Appl Phys. Lett.* 17, 257 (1970).
4. K. M. Evenson, G. W. Day, J. S. Wells and L. O. Mullen, *Appl. Phys. Lett.* 22, (1972).
5. A. Sanchez, S. K. Singh and A. Javan, *Appl. Phys. Lett.* 21, 240 (1972).
6. P. L. Richards, *Infrared Detectors II, Semiconductors and Semimetals*, R. K. Willardson and A. C. Beer, Eds. (Academic Press, New York 1975) Vol. 11 (to be published).
7. H. K. Fetterman et al., *IEEE Trans. MTT-22*, 1013 (1974).
8. A. H. Silver, *IEET Trans. MAG-11*, 794 (1975).
9. S. Silver, *Microwave Antenna Theory and Design*, (McGraw Hill, New York, 1949).
10. D. C. Sinclair and W. E. Bell, *Gas Laser Technology* (Holt, Reinhart and Winston, New York, 1969) Ch. 4.
11. J. G. Small, G. M. Elchinger, A. Javan, A. Sanchez, F. J. Bachner, and D. L. Smythe, *Appl. Phys. Lett.* 24, 275 (1974).
12. L. M. Matarrese and K. M. Evenson
13. B. Twu and S. E. Schwarz, *Appl. Phys. Lett.* 26 672 (1975).
14. J. P. Keller, *J. Appl. Phys.* 30 1452 (1959).
15. K. K. Chan, L. B. Felsen, A. Hessel and J. Shmoys, Paper VI 1.1, International Union of Radio Science Meeting, Urbana, Illinois, June 1972.

16. T. G. Blaney, Rev. Phys. Appliquee 9, 279 (1974).
17. K. Mizuno, R. Kuwahara and S. Ono, Appl. Phys. Lett. 26, 605 (1975).
18. C. W. Harrison, Jr., J. Appl. Phys. 14, 537 (1943).
19. H. Jasik, ed. Antenna Engineering Handbook, (McGraw-Hill Book Co., Inc., New York, 1961).

DEGRADATION OF PHOTOVOLTAIC CELLS

W. Kohn

ABSTRACT

This is a short report on a one-half day workshop on degradation of CdS/Cu_xS and similar photo-cells. For these systems, which can be very inexpensively prepared, the short operating life-time is the chief obstacle to large scale use. The present status of understanding of the degradation mechanism is briefly reviewed. It is noted that very little is firmly established and some suggestions for further research are made.

DEGRADATION OF PHOTOVOLTAIC CELLS

W. Kohn

I. BACKGROUND

As a result of the decreasing availability of oil as an energy source, there has been greatly increased interest in solar energy technologies, including photovoltaic cells. In 1973 and 1974 MRC organized meetings and studies dealing with solar energy reported in Volume II, 1973, and Volume II, 1974.

Currently most of the national R and D emphasis in this field is on high grade homojunction Si devices which are very attractive from all standpoints (materials availability, efficiency, stability) except, unfortunately, cost which is still about 2 orders of magnitude too high. A promising recent development in Si-technology is the new edgefed growth process being developed by Tyco. Other high efficiency, but expensive systems which have recently been reported are: InP/CdS at Bell Labs. (J. Shay) and GaAs/GaAlAs at Varian (L. James and R. Moon); the latter system has an excellent efficiency (~20%) and can operate at very high temperature, permitting its use in conjunction with strong light concentration devices.

A system whose photovoltaic properties have been known

for several decades is p- $\text{Cu}_x\text{S}/\text{n-CdS}$ ($x \approx 2$). This can be produced extremely cheaply (projected estimate $\sim \$10/\text{m}^2$) by dipping a thin CdS polycrystalline film into a warm CuCl solution. Such cells have initial efficiencies in the range of 5-8%. Generally, however, they degrade substantially over a period of weeks or months. Diffusion of Cu-ions is believed to play a major role in this process. This problem has been a key obstacle to further development of these cells.¹

Under these circumstances and because of the possible interest of one or more University groups in newly embarking on a pertinent research program, Professor R. H. Bube of Stanford University, one of the leading experts in this field, was invited as a consultant to MRC on July 8, 1975, and led a discussion of this subject area, parts of which are summarized in the following section. Responsibility for errors or inaccuracies in this account rests with the writer (W. K.). Other participants in the discussion included A. Bienenstock, H. Ehrenreich, R. Gomer, R. Huggins, E. Montroll, H. Reiss and A. Yariv.

II. THE NATURE OF THE DEGRADATION PROBLEM

It is presently believed that there are two main reasons for the degradation of $\text{Cu}_x\text{S}/\text{CdS}$ cells.

a. Oxygen Attack on the Cu_xS face

In typical operation the thin ($< 1\mu$) face of Cu_xS is exposed to light and to air. This results in a partial

oxidation of the top layers and to a reduction of the Cu content in the underlying Cu_xS . Compounds of the Cu_xS type exist over a rather wide composition range with $1.8 \leq x \leq 2.0$ in a variety of crystal structures. However, very remarkably, they are photovoltaically effective only in a very narrow range near $x \approx 1.9995$ (with presumably 0.0005 Cu vacancies).

An unpublished report by K. Bogus, European Space Research and Technology Center, Noordwijk, Belgium, states that the oxygen problem can be overcome by depositing a thin film of Cu on the Cu_xS face, leading to stable cells with an 8% efficiency.

b. Diffusion of Cu^+ into CdS Side

Cu^+ will diffuse into the CdS side of the junction where it gives rise to impurity states which change the nature of the electrical barrier. This problem was stated to have been overcome by W. Palz, Centre National d' Etudes Spatiales, France (~1973), by an unspecified method. It is conceivable that this was achieved by counterdoping the CdS (i.e., adding to the normal donors an additional number of donors and compensating acceptors); the acceptors may have an inhibiting effect on the diffusion. There is no indication that this work is continuing. (Why not?)

III. OTHER MISCELLANEOUS COMMENTS

An EPRI (Electric Power Research Institute) report on thin film CdS/ Cu_xS junctions is due in the near future.

A major program on this system is underway at the University of Delaware (A. Rothwarf, Karl Boer + (10-30) people).

There has been for a long time an R and D program at Westinghouse (F. A. Shirland).

A method of producing the Cu_xS layer on single crystals was developed at Philips, Eindhoven (Dieleman, Teveldi, Mulder), which is said to lead to stable 8% cells. It is believed that this work is not continuing. (Why not?)

R. M. Bube at Stanford is now getting away from systems involving the apparently troublesome Cu and studying other systems of the type p II-VI/n II-VI:

p		n	
Material	Gap (eV)	Material	Gap (eV)
CdTe	1.4	CdS	2.4
ZnTe	2.0	ZnSe	2.7

Light is generally absorbed on the p-side.

A problem area is lattice mismatch. Ideally this should be held to 0.1% (possible in III/V alloy systems), however in many systems one can still operate with mismatches as high as 4%.

A list of recent reports on Cu_xS/CdS is attached.

IV. POSSIBLE DIRECTIONS OF FUTURE RESEARCH

Future research should deal comparatively with single crystal CdS and film CdS systems. The former are more easily

reproduced and characterized, but only the latter are cheap enough to be of economic interest.

Diffusion studies of Cu^+ in CdS a function of temperature, light level, impurities, dislocations, grain boundaries.

Careful studies of the $\text{Cu}_x\text{S}/\text{CuO}$ system with special attention to obtaining and maintaining the ideal composition, $x = 1.9995$.

Detailed diagnostic measurement on junctions including Auger or ESCA measurements combined with gradual layer removal by Ar ion bombardment.

Determination of the activation energy for degradation. If this energy is high (≥ 0.5 eV), a technically interesting solution might be to use the cell not only photovoltaically but also photothermally. This would, in addition to the production of electrical energy, also provide a substantial amount of low temperature heat at $\leq 100^\circ\text{C}$. At the same time, removal of this heat would lead to a temperature reduction and hence to an increased life-time of the cell.

Finally, comparative studies of related systems, selected to be photovoltaically promising (direct band gap, band gaps of the right order, 1-2.5 eV, etc.), such as now conducted by Bube, are promising.

ACKNOWLEDGEMENT

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

Reference

1. P. Rappaport, MRC Report 1974, Vol. II, p. 170.

RECENT REPORTS ON Cu₂S/CdS CELLS

"Thin Film Solar Cells for Terrestrial Applications"
F.A.Shirland, W.J.Biter, E.W.Greenreich and T.P.Brody
NSF/RANN/SE/AER74-14918/TR/74/4 - Annual Progress Report

"Direct Solar Energy Conversion for Large Scale Terrestrial Use"
Institute of Energy Conversion, University of Delaware
NSF/RANN/SE/GI-34872/PR/73/4 - Annual Progress Report

"Proposed Definition of AM1 Solar Simulation For Standardized Testing of Solar Cells"
Institute of Energy Conversion, University of Delaware - H.C.Hadley, Jr.
NSF/RANN/SE/GI34872/TR74/3

"Direct Solar Energy Conversion for Large Scale Terrestrial Use"
Institute of Energy Conversion, University of Delaware
NSF/RANN/SE/GI-34872/PR74/2 - Semi-Annual Progress Report
NSF/RANN/SE/GI-34872/PR74/3 - Quarterly Progress Report
NSF/RANN/SE/GI-34872/PR74/4 - Annual Progress Report
NSF/RANN/AER72-03478 A03/PR75/1 - Quarterly Progress Report

"Electron Diffraction Study of Phases in the Surface Layer of CdS/Cu₂S Solar Cells"
Institute of Energy Conversion, University of Delaware - W.F.Tseng and I.G.Greenfield
NSF/RANN/SE/GI 34872/TR74/2

"Performance of CdS/Cu₂S Solar Cells Under Roof Top Insolation Conditions"
Institute of Energy Conversion, University of Delaware - H.M.Windawi
NSF/RANN/AER 72-03478 A03/TR 75/1

"Photovoltaics in Heterojunctions - Particularly CdS/Cu₂S"
Institute of Energy Conversion, University of Delaware - A.Rothwarf
NSF/RANN/AER 72-03478 A03/TR 75/4

THEORETICAL ISSUES RELATED TO SOLIDS
CONTAINING HIGHLY MOBILE ATOMS OR IONS

R. A. Huggins

ABSTRACT

Solids containing highly mobile atomic species have received greatly increasing attention recently, due at least partly to their potential practical use in battery and fuel cell systems, or as electrochemical transducers in a variety of other types of applications.

Theoretical understanding in this area is currently quite immature. As a guide to the development of useful theoretical efforts in this area, some of the important experimental facts that distinguish these materials are reviewed.

Similarities between fast ionic conductors and interstitial diffusion of hydrogen in certain metals will be shown. Misconceptions concerning the existence of "magic ions" and a "superionic-transition" are also discussed.

THEORETICAL ISSUES RELATED TO SOLIDS
CONTAINING HIGHLY MOBILE ATOMS OR IONS

R. A. Huggins

I. INTRODUCTION

There is currently a rapidly growing interest in various aspects of the science and technology of solids in which certain ions exhibit unusually rapid transport behavior. These materials, which are variously called fast ionic conductors, solid electrolytes, and superionic conductors have an unusual set of properties, directly related to their large values of diffusivity and ionic conductivity. In some cases, this rapid ionic motion is accompanied by appreciable, or even preponderant, electronic conduction. Materials with significant charge transport due to the motion of both ionic and electronic species are known as mixed conductors. Both ionic conductors and mixed conductors are of potential use in connection with a number of possible technological applications.

Because of the rapidly growing interest in this field among people with backgrounds in other areas, it is important that theoretical and experimental efforts focus upon essential features of the various problems in this area. It is the purpose of this paper to provide some information about experimental facts and current issues, with the hope that it may

serve as useful guidance to people entering this rapidly moving field.

Observation of a number of recent papers relating to fast ion conduction has pointed out the special need for this kind of guidance among those interested in undertaking meaningful theoretical work in this area. Since the scientific community expects theoretical efforts to contribute coherence, understanding and guidance in any field, it is imperative that recognition be given to the current state of understanding that has resulted from experimental efforts. Unfortunately, there is evidence that this is not always the case, however.

II. PRIMARY DIFFERENCES BETWEEN FAST IONIC CONDUCTORS AND

"NORMAL" IONIC CONDUCTORS

A. Conductivity Parameters

Measurements of ionic conductivity and diffusion on a large number of materials, predominantly the halides, over a period of many years has led to the establishment of a standard theoretical approach to the point defect structure of ionic solids, as well as to hopping-type models for mass and charge transport in such materials. Experimental results have shown that the transport behavior is dependent upon the disorder type, which indicates the intrinsic defect pair of lowest free energy. It is now evident that materials with Frenkel disorder, and thus interstitial species, typically have much greater values of ionic conductivity than those with Schottky disorder,

where the predominant defects are vacancies.

Quite early, it was also recognized that a number of these simple ionic compounds undergo phase transformations at elevated temperatures, with the ionic conductivity going through a corresponding abrupt transition. Much higher values were sometimes found in the high temperature phases, with a corresponding reduction in the temperature dependence.

As an example, the high conductivity of some silver salts at elevated temperatures was studied as early as 1912 by Tubandt and Lorenz¹. These same authors showed² that in the case of AgI, the conductivity in the solid state near the melting point is about 20% greater than it is in the liquid, in contradistinction to the traditional concept that solids have quite rigid structures, and hence relatively immobile ions.

Although α -AgI was the first example given much attention, other materials in which this type of behavior was recognized quite early were α -Li₂SO₄³, α -CuI⁴, α - and β -CuBr⁵, and α -Ag₂HgI₄⁶. Discussion of these early results was presented in a series of reviews⁷ by Tubandt and his co-workers. It should also be pointed out that, although these materials are all predominantly ionic conductors, it was also recognized quite early that ionic motion in some materials, such as α -Ag₂S and α -Cu₂Se⁸, which are mixed conductors, is also very rapid. There was some confusion about the identity

of the predominant charge carrier in the former for a short time^{9,10}.

Some attention was given as early as the 1920's¹¹⁻¹⁴ to the reasons for this unusual type of behavior, which is clearly different from that of the more "normal" Schottky and Frenkel disorder materials, and more like that typically associated with liquids.

A great wealth of experimental information has become available since that time, and it is now known that this type of "high temperature" behavior, characterized by unusually rapid ionic motion, can persist to ambient temperatures and below in some materials, and that it is not necessarily related to the existence of a first order phase transformation. Instead, there may be fundamental differences in the transport behavior of this third group of materials from the Schottky and Frenkel disorder types.

The difference between these three classes of ionic conductors can best be seen from the temperature dependence of their ionic conductivities.

It is generally found that in all three cases, the variation in the conductivity σ with temperature can be represented by an expression of the Arrhenius type,

$$\sigma = \left(\frac{\sigma_0}{T} \right) \exp (-\Delta H/RT) \quad (1)$$

Data indicate that there are three principal features in the conductivity which differentiate the fast ionic con-

ductor group from the others. The most obvious difference is in the absolute magnitude of the ionic conductivity, which can be unusually high over a wide range of temperature in some cases. In addition to the unusual magnitude of the ionic conductivity, the temperature dependence, represented by the activation enthalpy ΔH , is considerably lower in the fast ionic conductors. The value of the preexponential constant σ_0 , related to the intercept when $(1/T)$ approaches zero, is also considerably lower. These characteristics are obviously interrelated, rather than independent. Together, they clearly indicate that the process of ionic transport in the fast ionic conductor group is quite different from that in the other two.

There is also a fourth group of materials, including a number of halides with the fluorite (CaF_2) structure, which have anti-Frenkel disorder, which have relatively high activation enthalpies at lower temperatures, but which go through a gradual transition to fast ionic type of conduction (lower ΔH and σ_0) as their conductivity values become very large at high temperatures.

B. Thermal Properties and Sublattice Disorder

Another macroscopic indication of the difference between the fast ionic conductor group and the Schottky and Frenkel disorder groups lies in their thermal properties. As early as 1935 it was found¹⁵ that the specific heat in the β (low temperature) phase of Ag_2HgI_4 , near the β - α transition undergoes an abnormal rise, and that the enthalpy change at

the transformation is unusually low. This was interpreted as a disordering of the ions upon a dilutely populated set of essentially identical sublattice sites. The ionic conductivity also rises appreciably in this temperature range. X-ray diffraction experiments¹⁶ also had indicated that the high temperature α phase of this material (which behaves as a fast ionic conductor) has an "averaged" or disordered arrangement of ions upon the cation sublattice, with three ions distributed at random upon four lattice sites.

A similar conclusion about disordered occupancy of a dilutely populated set of cationic sites had also been arrived at by Strock^{17,18} for the case of α -AgI. Both Strock and Ketelaar^{6,19} recognized the connection between this microscopic structural feature and the unusually high ionic conductivity of these materials, although Strock¹⁸ who introduced the concept of a molten sublattice, contended that there was a significant difference between the two cases.

The existence of unusually large values of specific heat in cation-disordered phases was emphasized by Wiedersich and Geller²⁰ who pointed out that there are several cases in which the heat capacity at constant pressure decreases with increasing temperature over extensive ranges. Such measurements were reported by Johnston et al.²¹ for the fast ionic conductor RbAg_4I_5 . Thermal effects within the α -AgI phase, possibly due to changes in the order parameter, have also been reported²²⁻²⁴.

The existence of small entropy changes upon melting in fast ionic conductors, as well as large entropy changes associated with first order phase transformations in cases in which one phase (generally the one stable at the higher temperature) is a fast ionic conductor, and the other is not, has been pointed out by O'Keeffe²⁵⁻²⁷ and van Gool²⁸. O'Keeffe has also emphasized thermal evidence for order-disorder reactions in a number of anionic conductors, mentioned earlier as a fourth group in which there is a transition from behavior typical of the Frenkel-disorder group to properties characteristic of the fast ionic conductor class. From the magnitudes of the thermodynamic data, he suggests that the mobile ions in materials which are fast ionic conductors should be thought of as being in a liquid state, as proposed earlier for α -AgI by Strock¹⁷ on the basis of X-ray diffraction results.

It should be pointed out that thermal effects have been observed in a number of cases, e.g., Ag_2HgI_4 ¹⁵, Cu_2HgI_4 and CuI ²⁹, which indicate a high degree of disorder in phases at temperatures below first order phase transformations. Disordered sublattices are obviously not restricted to high temperature phases.

C. X-ray Studies

Following the early X-ray diffraction work by Strock^{17, 18} and Ketelaar¹⁶, which showed a large measure of disorder in materials then recognized as fast ionic conductors, other investigators have used more refined techniques to study the

details of this phenomenon from a structural point of view.

Some years ago, attention was given to the structure of the copper sublattice in "high chalcocite", $\alpha\text{-Cu}_2\text{S}$, which is a mixed conductor. It transforms to an ordered "low chalcocite" upon cooling at about 105°C. It was first concluded^{30, 31} that the cation sublattice was completely disordered in the higher temperature phase. Careful work by Buerger and Wuensch³², however, showed that random distribution of atoms over the equipoints was not correct. Refinement of the data showed that the electron density related to the copper sublattice was appreciably smeared out, extending between the three different types of possible copper sites derived from the geometry of the hexagonal close-packed sulfur lattice.

They concluded that, even if anisotropic temperature factors were used, placement of discrete atoms at the sites of 2-fold, 3-fold, and 4-fold symmetry would not be a satisfactory description of the structure, for the data clearly indicated a continuous copper distribution between these possible sites.

Work on powdered samples of $\alpha\text{-CuI}$ and $\alpha\text{-CuBr}$ by Miyake and his collaborators^{33, 34} showed an anomalous increase in the intensities of certain reflections with temperature. This was interpreted as being due to either a greater Debye temperature factor for cations than for anions, or an array of four possible minimum-energy positions displaced in the $\langle 111 \rangle$ directions from the center of symmetry of the normally assumed copper sites, with the occupancy probability varying with temperature. Such

an explanation was found to fit the entropy data reasonably well. Hoshino³⁵ also reported similar conclusions from a study of polycrystalline Ag_2HgI_4 , and remarked that another possible interpretation might involve asymmetric anharmonic thermal vibration of the cations, with greater amplitudes in the <111> directions.

Single crystal X-ray diffraction experiments have been recently reported on $\alpha\text{-Ag}_2\text{HgI}_4$ by Kasper and Browall³⁶. They found that one could fit the data adequately by the Ketelaar¹⁶ model, but this required the use of large thermal parameters. However, better fit was obtained by assuming Hoshino's³⁵ static displacement model. Comparable fit can also be obtained by analysis in terms of anharmonic vibrational motion, as has been found to be present in other crystals (e.g., CaF_2) containing ions at or near sites of $\overline{43}\text{ m}$ symmetry³⁷⁻³⁹ in which tetrahedral displacement effects are indicated. The theory of anharmonic effects, and its application to zincblende structures was presented by Cooper et al.⁴⁰, and Kasper and Browall found that their data fit the necessary criteria quite well.

Thus, it appears that in at least some fast ionic conductors it is quite reasonable to think in terms of the mobile cations having unusually large vibrational amplitudes in preferred directions, with a considerable degree of anharmonic character.

D. Structure-Dependence of Fast Ionic Conduction

One of the very important factors determining which

materials are fast ionic conductors is the configuration of the relatively static sublattice through which the mobile species move. This phenomenon is extremely structure-dependent, and it has become recognized that many of the known fast ionic conductors have static sublattice structures of certain types. If modelled by using hard spheres of appropriate radii to represent the constituent ions, they are found to always contain pathways of connected space (crystallographic tunnels) with lateral dimensions comparable to the diameters of the mobile species.

It is obvious that because of this geometric feature of their crystal structures, the motion of ions throughout the structure can occur without significant ionic displacements and large strain energy effects, such as are found to dominate the activation enthalpy in close-packed metallic systems. In favorable cases, the various other terms in the potential energy also tend to balance each other out, leading to unusually flat potential energy profiles along the preferred transport path for the mobile species^{41,42}.

The categorization of fast ionic conductors in terms of the structural arrangement of the static sublattice has been presented in several places⁴³⁻⁴⁵. A number of examples are included in Table I.

TABLE I. CATEGORIZATION OF SOME FAST IONIC AND MIXED CONDUCTORS
ACCORDING TO STRUCTURE OF STATIC SUBLATTICE

<u>Body-Centered Cubic Structure</u>		<u>Layer Structures</u>
α -AgI	Ag ₃ S Br	β -aluminas
α -Ag ₂ S	(Li, Na) ₂ SO ₄	β'' -aluminas
α -Ag ₂ Se	(Li, Ag) ₂ SO ₄	silver organic iodides
α -CuBr		transition metal dichalcogenides
α -Ag ₃ SI		graphite compounds
β -Ag ₃ SI		
<u>Face-Centered Cubic Structure</u>		<u>Hexagonal Structure</u>
α -CuI	α -Ag ₂ Te	Na ₂ SO ₄
α -Cu ₂ S	α -Ag ₂ HgI ₄	Ag ₂ SO ₄
α -Cu ₂ Se	α -Cu ₂ HgI ₄	K ₂ SO ₄
α -Li ₂ SO ₄	α -(Ag, Cu)HgI ₄	LaF ₃
α -Li ₂ WO ₄		
<u>Other Linear Tunnel Structures</u>		
tetragonal tungsten bronzes		
hexagonal tungsten bronzes		
hollandites		
ramsdellites		

III. ANALOGOUS BEHAVIOR OF INTERSTITIAL SPECIES IN SOME METALS

It has long been known that certain elements diffuse with unusual rapidity in some metals. The primary example of this is the transport of hydrogen through palladium, which has led to the technological use of palladium as a hydrogen-selective membrane.

It is now apparent that several of the characteristics of the diffusion of hydrogen in metals are quite analogous to the very rapid transport of certain species in ionic crystals. As in the latter case, the crystal structure of the "host" lattice is of primary importance. From available data, one sees that hydrogen diffuses in body-centered cubic metals very rapidly at unusually low temperatures, and the parameters in the Arrhenius equation that describe the experimental data are very similar to those found in fast ionic conductors. The pre-exponential factor D_0 and the activation enthalpy ΔH are both unusually low. Available experimental data on hydrogen diffusion in metals were reviewed by Birnbaum and Wert⁴⁶ and more recently by Völkl and Alefeld⁴⁷.

Although hydrogen also moves rapidly through metals with the face-centered cubic structure, the parameters for diffusion in such cases are somewhat different from those in body-centered cubic metals, the preexponential and the activation enthalpy both being greater in the face-centered cubic case.

Table II shows the parameters in the Arrhenius relation $D = D_0 \exp(-\Delta H/RT)$ relating to the diffusion of hydrogen in a group of body-centered metals, as well as in a few important face-centered cubic metals. These data were taken from the compilation by Völkl and Alefeld⁴⁷. It can be seen that the diffusion coefficient at room temperature is not greatest in palladium, a face-centered cubic metal. Nonetheless, it is used instead of the body-centered cubic metals as a practical hydrogen-selective membrane because it does not readily oxidize, as they do. Such oxide films interfere with the permeation of hydrogen. It should also be pointed out that, although platinum and palladium have comparable diffusion coefficients, the solubility of hydrogen in palladium is much greater, so that the rate of permeation, which is proportional to the product of the solubility and the diffusion coefficient, is considerably greater in palladium than in platinum.

Because of the unusual transport parameters that are observed, Alefeld⁴⁸ has argued that hydrogen in body-centered cubic metals should be thought of as having the characteristics of a liquid. This is, of course, analogous to the molten sub-lattice model of fast ionic conductors.

IV. FAST DIFFUSION IN SEMICONDUCTORS

Experimental data indicate that diffusion behavior of solute species in the common group IV semiconductors, and probably also in the III-V compounds, falls into two quite

TABLE II. PARAMETERS FOR HYDROGEN DIFFUSION IN METALS

B C C METALS

	D_0	ΔH (kJ/Mole)	D at $25^\circ C$ (cm^2/sec)
V	2.9×10^{-4}	4.15	5×10^{-5}
Nb	5.0×10^{-4}	10.2	8×10^{-6}
Ta	4.4×10^{-4}	13.5	2×10^{-6}
Fe	4.0×10^{-4}	4.5	8×10^{-5}
	7.5×10^{-4}	8.5	1×10^{-5}

F C C METALS

Pt	6.0×10^{-3}	25.1	3×10^{-7}
Pd	2.9×10^{-3}	22.2	3×10^{-7}
Ni	4.8×10^{-3}	39.6	5×10^{-10}

disparate general categories. Solutes are often classified as "slow diffusers" or "fast diffusers". Among the elements generally accepted as fast diffusers in germanium are Li, Cu, Ag, Fe, Co, and Ni. Likewise, fast diffusion in silicon is experienced by Li, Na, K, Cu, Ag, Au, Fe, and Ni. This type of behavior was noticed in the early days of the development of semiconductor technology^{49,50}, as rapid diffusion of some species was causing great practical difficulties. A review of the early work in this area was presented by Letaw⁵¹. It was found that the fast diffusing species was interstitial copper, and it was recognized by van der Maesen and Brenkman⁵² that copper atoms may exist on both substitutional and interstitial sites in these diamond structure semiconductors. The interstitial species are very mobile and the total concentration profile involves the interaction of interstitial and substitutional species. Frank and Turnbull⁵³ showed that the role of vacancies must also be taken into account in this type of transport process, which occurs by what is now commonly known as the "dissociative mechanism".

Diffusion data on various species in germanium and silicon were summarized by Seeger and Chik⁵⁴. It is generally found that the D_0 values for these interstitial diffusers in germanium and silicon are similar to those for hydrogen in metals (in the range of $1-6 \times 10^{-3} \text{ cm}^2/\text{sec}$). The activation enthalpies, while quite a bit greater than for fast ionic or hydrogen diffusion, are also relatively low (from 29 to over

97 kJ/mole) in contrast with the values found for diffusion of the group III and V elements in germanium and silicon, which tend to be between 300 and 400 kJ/mole. The preexponential factors are also much higher in the latter cases.

V. FAST IONIC CONDUCTION IN GLASSY AND TWO-PHASE STRUCTURES

Although it is generally found⁵⁵ that glasses have modest values of ionic conductivity, it has been reported in some cases that materials with a disordered structure have unusually high ionic conductivity. An example was the observation⁵⁶ of large Ag⁺ conductivity in disordered samples of AgI/Ag₂SeO₄ at ambient temperatures. In addition, it was recently found⁵⁷ that lithium ion conduction is more rapid in disordered samples than in more crystalline samples of the lithium aluminosilicate beta eucryptite.

It also seems that the presence of second phase particles of supposedly inert and insoluble phases plays a role in enhancing the lithium ionic conductivity in polycrystalline LiI at ambient temperatures⁵⁸.

Neither of these observations has been satisfactorily explained.

VI. SOLID SOLUTION EFFECTS

It has been known from the work of C. Wagner and his associates in the 1930's that the conductivity of "normal" ionic conductors can be increased by the enhancement of the concentration of native defects by the necessity to balance

the electrostatic charge of aliovalent dopants. The same thing is sometimes found in some fast ionic conductors. For example, divalent cation solutes, such as Ca^{+2} or Mg^{+2} greatly increase the conductivity of the elevated temperature hexagonal phase of Na_2SO_4 ⁵⁹. However, the conductivity of Li_2SO_4 is decreased, rather than increased, by the presence of Mg^{+2} or Ba^{+2} ⁶⁰.

Other puzzling results have been reported on solid solutions of sulfates by the group at the Chalmers Technical University at Goteborg. For example, isovalent cation solutes, which are not expected to influence defect concentrations in the normal sense, sometimes increase the conductivity (Li^+ in Na_2SO_4 ⁵⁹, K^+ and Rb^+ in Li_2SO_4 ⁶¹), and sometimes decrease it (Na^+ in Li_2SO_4 ⁶¹).

Anionic doping, such as occurs when LiCl is added to Li_2SO_4 , was found⁶⁰ to increase the conductivity, as expected.

In some cases, such as the sulfates, fast ionic conducting phases can exist over wide ranges of composition. It has also been found that a second phase with a different structure can be formed that is also a fast ionic conductor (e.g., in the $\text{Li}_2\text{SO}_4/\text{Na}_2\text{SO}_4$ ⁶², $\text{Li}_2\text{SO}_4/\text{Ag}_2\text{SO}_4$ ⁶³, and $\text{Li}_2\text{SO}_4/\text{ZnSO}_4$ ⁶⁴ systems).

VII. IONIC SIZE EFFECTS

It has been found that ion exchange can be accomplished in some cases, allowing the observation of the relative behavior of different mobile ions in essentially the same static structure.

Yao and Kummer⁶⁵ showed that the mobile ions in the bridging layer of the beta alumina structure could be exchanged with a number of other monovalent cations by immersion in appropriate molten salts. It has also been shown⁶⁶ that this can be done by electrochemical pumping.

Experiments using radiotracer techniques for determining the self diffusion coefficients of several different monovalent (M^+) ions in the beta alumina structure in the temperature range 200-400°C were reported by Yao and Kummer⁶⁵ and Radzilowski et al.⁶⁷. The dielectric loss method was also used to evaluate ionic motion in polycrystalline β alumina containing various different M^+ ions by Radzilowski et al.⁶⁷.

Using tungsten and vanadium bronze phases as electrodes, reliable data have been obtained for the ionic conductivity of single crystals of beta alumina containing sodium⁶⁸, potassium, thallium and lithium⁶⁹ from over 800°C to well below ambient temperature. Comparable results were obtained⁷⁰ on samples containing silver ions by the use of silver electrodes. Arrhenius plots of these measurements of the conductivity are linear over this whole temperature range.

The information acquired by the use of these three methods, radiotracer self diffusion, dielectric loss, and ionic conductivity, was found to be quite consistent over an extremely wide temperature range, and comparisons of the data for beta alumina containing sodium silver, potassium thallium, and lithium in the bridging layer have been presented^{44, 69}. These

transport measurements showed that there is an optimum ionic radius for achieving the maximum mobility in the beta alumina structure, with both smaller and larger ions less mobile than those of intermediate size. The data indicated that σ_0 is essentially independent of the identity of the M^+ ion, the major difference being due to variations in the activation enthalpy.

The importance of this size factor was convincingly confirmed by experiments by Radzilowski and Kummer⁷¹ which compared the influence of hydrostatic pressure upon the resistivity of beta alumina containing lithium, sodium and potassium ions. At higher pressures the resistivity of lithium β decreased, whereas it increased for potassium β , and was essentially unchanged in the sodium case. Further experiments confirming these results and going to higher pressures have recently been performed in Japan.

This type of ionic size dependence was predicted in the α -AgI structure by a structurally explicit calculation⁴¹ based upon the Born-Mayer method.

VIII. FURTHER UNUSUAL EXPERIMENTAL OBSERVATIONS

A. Enhanced Conductivity at Very High Temperatures

There have been a number of cases in which it has been experimentally observed that there is a positive deviation from the expected Arrhenius type of temperature dependence of the conductivity at high temperatures in some materials in which atoms or ions are already quite mobile, and presumably fully disordered.

This has been observed in the case of silver ion conduction in silver beta alumina⁷⁰. Positive deviations from Arrhenius behavior have also been observed in measurements of hydrogen diffusion in palladium⁴⁷. In addition, this type of behavior has been observed in several silver and copper salts which have Frenkel disorder⁷².

No satisfactory explanation has yet been presented for these observations, although they may indicate the presence of a second transport mechanism in this temperature range, such as the direct interstitial mechanism postulated⁶⁹ to explain the change in activation enthalpy with temperature in lithium beta alumina.

B. Conductivity Saturation

Deviations from Arrhenius behavior in the opposite direction have been found in a number of cases as well.

It has been found that the temperature dependence of the ionic conductivity decreases at high temperatures in some materials. Examples showing this effect include several anionic conductors with the fluorite structure. Materials with this structure tend to have anti-Frenkel disorder and high values of both the activation enthalpy and the preexponential factor. Thus, at temperatures near the melting point the actual conductivity values become quite high, of the order of magnitude of those found in molten salts. In this region the conductivity becomes more and more temperature-independent, the apparent activation enthalpy thereby tending toward zero.

This phenomenon is not restricted to fluorite structure materials. It is also found in Li_4SiO_4 ⁷³, and solid solutions based upon Na_2SO_4 ⁵⁹.

This kind of behavior implies that there is some maximum limit upon the conductivity, and when this limit is approached the conductivity exhibits temperature independence.

Although there have been several mechanistic interpretations of extremely high conductivity values that may relate to this apparent saturation effect^{25, 74}, this type of behavior deserves more attention from both theoretical and experimental viewpoints.

IX. SOME AREAS OF MISUNDERSTANDING OR CONFUSION

A. The "Magic Ion" Fallacy

It was thought for some time that there was something special about silver and copper ions, and that all fast ionic conductors contain one of those species. Although it is true that most of the solids which were found in the early days to be fast ionic conductors were indeed silver or copper salts, this is by no means an exclusive group, for it was found as early as 1921 that Li_2SO_4 is a fast ionic conductor for lithium ions at high temperatures.

At the present time quite a variety of materials are recognized as being fast ionic conductors. Included are not just silver and copper salts but also those containing lithium, sodium, and potassium among the cation group. Fast anionic

conductors include fluorides, chlorides, and oxides.

Some of those materials which are known to be fast ionic conductors are listed in Table II.

This again reinforces the contention that the important feature leading to fast ionic conductivity is the crystal structure rather than the identity of the specific species present upon the lattice.

B. The "Superionic Transition" Fallacy

Unfortunately, there seems to be some confusion among a number of people now working in this area about the roles of both first order and second order phase transitions in some of the materials which show rapid ionic motion.

Indeed, the titles and content of some recent papers have implied that there is a "superionic transition", in some manner analogous to the electronic superconductive transition. This is quite unfortunate, as the physical phenomena involved in these processes are distinctly different.

This concept of a "transition" undoubtedly arose from the observation that there is an abrupt discontinuity in the ionic conductivity in some of the recognized fast ionic conductors, such as AgI. Unfortunately, this is not a unique characteristic of these materials, and it is not present in some of the more important examples, such as the beta alumina family. Actually, as will be mentioned shortly, materials which are fast ionic conductors show a wide range of phase transition behavior, as do solids in general - including first

TABLE III. EXAMPLES OF MATERIALS THAT ARE FAST IONIC OR MIXED CONDUCTORS

CATION CONDUCTORS

AgI	CuI	Li ₂ SO ₄	Na ₂ SO ₄
Ag ₂ S	CuBr	Li ₂ WO ₄	Na- β alumina
Ag ₃ Si	Cu ₂ S	Li _x TiS ₂	NaSbO ₃
Ag ₄ RbI ₅	Cu ₂ HgI ₄	Li _x V ₂ O ₅	Na ₂ WO ₄
Ag- β alumina	Cu- β alumina	Li ₂ Ti ₃ O ₇	K- β alumina
AgBr	Cu ₂ Se	Li ₄ SiO ₄	KFe ₁₁ O ₁₇
Ag ₂ Se	Cu ₄ KI ₅	LiCuFe(CN) ₆	NaGa ₁₁ O ₁₇
Ag ₃ SBr		Li- β alumina	K ₂ SO ₄
Ag ₂ HgI ₄		(Li,Na) ₂ SO ₄	KHF ₂
Ag ₂ Te		(Li,Ag) ₂ SO ₄	
Ag ₄ KI ₅		Li _x Pb _y SO ₄	
AgSbO ₃		Li ₄ GeO ₄	
Ag ₂ SO ₄			

ANION CONDUCTORS

LaF ₃	SrCl ₂	CeO ₂
LuF ₃		ZrO ₂
PbF ₂		ThO ₂
CaF ₂		Bi ₂ O ₃

order transitions, second order transitions, congruent melting points, peritectic reactions, wide stoichiometric ranges, eutectoid reactions, off-stoichiometric compositions, or no discernible transitions at all.

In some cases such phenomena exert a large influence upon the ionic conductivity, in other cases the effects are minor, or nonexistent.

It is quite reasonable to assume that any material having a disordered sublattice at elevated temperatures might exhibit an order-disorder transition upon cooling due to the larger values of both configurational and vibrational entropy expected with the disordered sublattice, which could lead to the intersection of the free energy curves for these two structures as the temperature is reduced. This has indeed been found in a number of cases. Examples are the alkaline earth dihalides; it is even seen in beta alumina at low temperatures.

Interestingly, however, it is found that there is apparently no sharp change in the magnitude of the ionic conductivity as a result of such a sublattice ordering reaction, although in some cases, a change in the temperature dependence of the conductivity is apparent. In general, the ordered structure has a higher activation enthalpy, and also a higher value of the preexponential factor.

On the other hand, some fast ionic conductors have structures which become unstable upon cooling, and thus show first-order transitions. There are some cases, such as CuBr

and Ag_3Si , in which both structures exhibit fast ionic conductivity, although there is a discontinuity in the actual conductivity value across the phase transformation temperature. There are a number of cases in which the high temperature phase is a fast ion conductor and the lower temperature phase, whose structure results from the transformation due to the instability upon cooling, has ionic conductivity of the more typical classical type.

There also are cases in which a first order phase transition upon cooling of a "normal" phase results in the formation of a fast ionic conductor phase below the transition (e.g., H in Fe). A further possibility, which has also been found (Na_2WO_4 , Na_2MoO_4)⁷⁵⁻⁷⁷, is an intermediate temperature phase which is a fast ionic conductor, whereas both the lower and higher temperature structures are not.

It is obvious that although some materials which are fast ionic conductors in certain temperature regions do undergo first order transitions upon cooling, this is most properly thought of as resulting from structural instability, and not an inherent characteristic of fast ionic conductors. Some of the most important fast ionic conductors, such as the beta aluminas, show no signs of any first-order transition.

It is abundantly evident that the ionic transport is characteristic of the structure of a given phase, and independent of the question of its stability relative to other possible phases.

From a practical standpoint, it may be desirable to try to shift the transformation temperature in cases in which one of the phases has more desirable properties than the other. The temperature of some first-order transformations can be shifted appreciably by the presence of certain solutes. An example is the influence of OH^- upon the rare earth trifluorides^{7,8}.

C. The High Defect Concentration Quandry

It has been pointed out by several authors that fast ionic conductors can be viewed as having an extremely large concentration of mobile species compared to the more traditional ionic conductors, as the dilute site occupancy infers that essentially all of the ions present are free to move. This has led to the temptation to attribute the large values of conductivity primarily to this high mobile species concentration. Unfortunately, however, such an approach is not consistent with the common observation of an unusually low value of the pre-exponential factor, as discussed in Section B-1. Since the preexponential factor is proportional to the concentration of mobile species, if nothing else were to change, an unusually large concentration should tend to make it especially large; just the opposite of what is found. Evidently, the influence of other factors, such as the unusually low vibrational (attempt) frequency, overwhelms the influence of the concentration factor.

X. SUMMARY

Solids with unusually large values of ionic conductivity and diffusion coefficients are attracting a substantial

amount of attention because of their special properties and the possibility of their use in a number of important applications. One indication of this is the number of recent symposia or topical meetings dealing with such matters⁷⁹.

Despite the large amount of current activity, there are still two fundamental questions that need to be answered:

- i) what are the physical mechanisms involved in charge and mass transport in fast ionic conductors, and are they fundamentally different from those involved in "normal" ionic conductors, i.e., those with Frenkel disorder?
- ii) why do some materials exhibit this type of behavior, and other not?

Useful clues concerning the most appropriate ways to deal with such questions can often be gleaned from available experimental results. It has been the purpose of this paper to point out some features of the observed behavior of these materials that might be of special interest to those undertaking relevant theoretical work.

It may also be useful to draw attention to a number of publications that have reviewed some of the work in this area⁸⁰.

XI. ACKNOWLEDGEMENT

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

REFERENCES

1. C. Tubandt and E. Lorenz, Nernst Festschrift, W. Knapp, Halle (1912), p. 446.
2. C. Tubandt and E. Lorenz, Z. Phys. Chem. 87, 513 (1914).
3. A. Benrath and K. Drekopf, Z. Phys. Chem. 99, 57 (1921).
4. C. Tubandt, H. Reinhold and W. Jost, Z. anorg. allgem. Chem. 177, 253 (1928).
5. J. Geiler, Dissertation, Halle (1928).
6. J. A. A. Ketelaar, Z. Phys. Chem. B26, 327 (1934).
7. C. Tubandt et al., in Landolt-Börnstein: Phys.-chem. Tabellen, Hw. II, p. 1062 ff., (1923), Ergänzungsband I, p. 582 ff. (1927), Ergänzungsband II, part 2, p. 1042 ff. (1931), Ergänzungsband III, part 3, p. 2011 ff. (1936), Berlin.
8. H. Reinhold and H. Moehring, Z. physik. Chem. B38, 221 (1937).
9. K. Hauffe, Reaktionen in und an Festen Stoffen, Springer-Verlag (1966), p. 69 ff.
10. C. Wagner, Z. phys. Chem. B21, 42 (1933).
11. A. Smekal, Z. Physik. 26, 707 (1925).
12. A. Smekal, Z. Phys. Chem. B6, 103 (1929).
13. W. Jost, Z. Phys. Chem. B6, 88, 210 (1929).
14. W. Jost, Z. Phys. Chem. B7, 234 (1930).
15. J. A. A. Ketelaar, Z. Phys. Chem. B30, 53 (1935).
16. J. A. A. Ketelaar, Z. Krist. 87, 436 (1934).
17. L. W. Strock, Z. Phys. Chem. B25, 441 (1934).
18. L. W. Strock, Z. Phys. Chem. B31, 132 (1936).
19. J. A. A. Ketelaar, Trans. Faraday Soc. 34, 874 (1938).
20. H. Wiedersich and S. Geller, in The Chemistry of Extended Defects in Non-Metallic Solids, Eds. L. Eyring and M. O'Keeffe, North-Holland Pub. Co., Amsterdam (1970), p. 629.

21. W. V. Johnston, H. Wiedersich and G. W. Lindberg, *J. Chem. Phys.* 51, 3739 (1969).
22. C. M. Perrott and N. H. Fletcher, *J. Chem. Phys.* 48, 2143 (1968).
23. C. M. Perrott and N. H. Fletcher, *J. Chem. Phys.* 48, 2681 (1968).
24. C. M. Perrott and N. H. Fletcher, *J. Chem. Phys.* 50, 2770 (1969).
25. M. O'Keeffe, in Fast Ion Transport in Solids, Ed. W. van Gool, North-Holland Pub. Co., Amsterdam (1973), p. 233.
26. M. O'Keeffe, *J. Electrochem. Soc.* 121, 102C (1974).
27. M. O'Keeffe, *Bull. Am. Phys. Soc.*, Series II, 20, 451 (1975).
28. W. van Gool, in Phase Transitions, Ed. by H. K. Henisch, R. Roy and L. E. Cross, Pergamon, New York (1973) pp. 373-377.
29. G. W. Herzog and H. Krischner, in *Proc. 7th Int. Symp. on Reactivity of Solids*, Eds. J. S. Anderson, M. W. Roberts and F. S. Stone, Chapman and Hall, London (1972), p. 140.
30. M. J. Bueger, *Anais Acad. Brasil. Ciene.* 21, 261 (1949).
31. R. Ueda, *J. Phys. Soc. Japan* 4, 287 (1949).
32. M. J. Buerger and B. J. Wuensch, *Science* 141, 276 (1963).
33. S. Miyake, S. Hoshino and T. Takenaka, *J. Phys. Soc. Japan* 7, 19 (1952).
34. S. Hoshino, *J. Phys. Soc. Japan* 7, 560 (1952).
35. S. Hoshino, *J. Phys. Soc. Japan* 10, 197 (1955).
36. J. S. Kasper and K. W. Browall, *J. Solid State Chem.* 13, 49 (1975).
37. B. T. M. Willis, *Acta Cryst.* 18, 75 (1965).
38. B. Dawson, A. C. Hurley and V. W. Maslen, *Proc. Roy. Soc. A298*, 289 (1967).
39. B. T. M. Willis, *Acta Cryst.* A25, 277 (1969).
40. M. J. Cooper, K. D. Rouse and H. Fuess, *Acta Cryst.* A29, 49 (1973).

41. W. H. Flygare and R. A. Huggins, *J. Phys. Chem. Solids.* 34, 1199 (1973).
42. O. B. Ajayi, L. E. Nagel, I. D. Raistrick and R. A. Huggins, *J. Phys. Chem. Solids* 37, 167 (1976).
43. L. Heyne, *Electrochim. Acta* 15, 1251 (1970).
44. R. A. Huggins, in Diffusion in Solids/Recent Developments, Eds. A. S. Nowick and J. J. Burton, Academic Press, New York, (1975), p. 445.
45. R. A. Huggins, "Ionically Conducting Solid State Membranes," to be published in *Adv. in Electrochem. and Electrochem. Eng.*, Vol. 10, Ed. C. Tobias and H. Gerischer, John Wiley (1976).
46. H. K. Birnbaum and C. A. Wert, *Ber. Bunsenges, Phys. Chem.* 76, 806 (1972).
47. J. Völkl and G. Alefeld, in Diffusion in Solids/Recent Developments, Eds. A. S. Nowick and J. J. Burton, Academic Press, New York (1975), p. 231.
48. G. Alefeld, *Ber. Bunsenges, Phys. Chem.* 76, 355 (1972).
49. H. C. Theurer and J. H. Scaff, *J. Metals* 4, 59 (1951).
50. C. S. Fuller, H. S. Theurer and W. van Roosbroeck, *Phys. Rev.* 85, 678 (1952).
51. H. Letaw, Jr., *J. Phys. Chem. Solids* 1, 100 (1956).
52. F. van der Maesen and J. A. Brenkman, *Philips Research Reports* 9, 225 (1954).
53. F. C. Frank and D. Turnbull, *Phys. Rev.* 104, 617 (1956).
54. A. Seeger and K. P. Chik. *Phys. Stat. Sol.* 29, 455 (1968).
55. H. J. Frischat, Ionic Conduction in Glasses, Springer-Verlag (1975).
56. D. Kunze, in Fast Ion Transport in Solids, Ed. W. van Gool, North-Holland (1973), p. 405.
57. R. T. Johnson, Jr., B. Morosin, M. L. Knotek, and R. M. Biefeld, presented at American Physical Society Meeting, Denver (1975).
58. C. R. Schlaikjer and C. C. Liang, in Fast Ion Transport in Solids, Ed. W. van Gool, North-Holland (1973), p. 685.

59. B. Heed, Dissertation, Chalmers (1975).
60. A. Bengtzelius, Dissertation, Chalmers (1973).
61. A. Kvist, Dissertation, Chalmers (1967).
62. T. Förland and J. Krogh-Moe, Acta Cryst. 11, 224 (1958).
63. H. Øye, Dissertation, Trondheim (1963).
64. K. Schroeder, Dissertation, Chalmers (1975).
65. Y. F. Y. Yao and J. T. Kummer, J. Inorg. Nucl. Chem. 29, 2453 (1976).
66. M. S. Whittingham, R. W. Helliwell and R. A. Huggins, U.S. Gov. Res. and Dev. Report 69, 158 (1969).
67. R. H. Radzilowski, Y. F. Yao and J. T. Kummer, J. Appl. Phys. 40, 4716 (1969).
68. M. S. Whittingham and R. A. Huggins, J. Chem. Phys. 54, 414 (1971).
69. M. S. Whittingham and R. A. Huggins, in Solid State Chemistry, Ed. R. S. Roth and S. J. Schneider, Nat. Bur. Standards Spec. Pub. 364, Washington (1972), p. 139.
70. M. S. Whittingham and R. A. Huggins, J. Electrochem. Soc. 118, 1 (1971).
71. R. H. Radzilowski and J. T. Kummer, J. Electrochem. Soc. 118, 714 (1971).
72. J. K. Aboagye and R. J. Friauf, Phys. Rev. B11, 1654 (1975).
73. A. R. West, J. Appl. Electrochem. 3, 327 (1973).
74. H. Rickert, in Fast Ion Transport in Solids, Ed. W. van Gool, North-Holland Pub. Co., Amsterdam (1973), p. 3.
75. G. H. J. Broers, Ph.D. Thesis, Amsterdam (1958).
76. P. H. Bottelberghs and F. R. van Buren, J. Solid State Chem. 13, 182 (1975).
77. P. H. Bottelberghs, Ph.D. Thesis, Utrecht (1976).
78. R. C. Pastor and M. Robinson, Mat. Res. Bull. 9, 569 (1974).
79. a) Meeting on Mass Transport in Non-metallic Solids, Basic Science Section of Brit. Ceram. Soc., London (1969).

- b) Symposium on Solid State Chemistry, Nat. Bur. Stand., Washington (1971).
- c) NATO Advanced Study Institute on Fast Ion Transport in Solids, Belgirate (1972).
- d) Session on Solid Electrolyte Fundamentals and Applications, Electrochem. Soc., Houston (1972).
- e) 24th Meeting of ISE, Solid Electrolytes, Ion Transport in Insulating Layers, and Their Application, Eindhoven (1973).
- f) Session on High Temperature Electron and Ion Transport in Solids, Electrochem. Soc., San Francisco (1974).
- g) 9th Univ. Conf. on Ceram. Science, Cleveland (1974).
- h) Session on Ceramics for Fuel Cell Applications, Am. Ceram. Soc., Williamsburg (1974).
- i) Symposium on Superionic Conductors, Solid State Physics Division of Am. Phys. Soc., Denver (1975).
- j) Session on Ceramic Materials for High Power Density Batteries, Am. Ceram. Soc., Washington (1975).
- k) Session on Materials Problems in Battery and Fuel Cell Technology, Electrochem. Soc., Toronto (1975).

80. a) D. O. Raleigh, "Solid State Electrochemistry" in Progress in Solid State Chemistry, Vol. 3. Ed. by H. Reiss, Pergamon (1967) p. 83.

b) L. Heyne, Electrochim. Acta 15, 1251 (1970).

c) B. B. Owens, Adv. Electrochem. Electrochem Eng. 8, 1 (1971).

d) J. T. Kummer, " β -Alumina Electrolytes", in Progress in Solid State Chemistry, Vol. 7. Ed. by H. Reiss and J. O. McCaldin, Pergamon (1972) p. 141.

e) B. C. H. Steele, "Electrical Conductivity in Ionic Solids", in Solid State Chemistry. Ed. by L. E. J. Roberts, Butterworths (1972) p. 117.

f) M. S. Whittingham and R. A. Huggins, "Beta Alumina-Prelude to a Revolution in Solid State Electrochemistry", in Solid State Chemistry. Ed. by R. S. Roth and S. J. Schneider, Nat. Bur. Standards Spec. Pub. 364 (1972) p. 139.

- g) J. Hladik, editor, Physics of Electrolytes, Vol. 1 and 2. Academic Press (1972).
- h) W. van Gool, editor, Fast Ion Transport in Solids, North-Holland (1973).
- i) W. van Gool, "Fast Ion Conduction", in Annual Review of Materials Science, Vol. 4. Ed. by R. A. Huggins, Annual Reviews, Inc. (1974) p. 311.
- j) R. A. Huggins, "Very Rapid Ionic Transport in Solids", in Diffusion in Solids: Recent Developments. Ed. by A. S. Nowick and J. J. Burton, Academic Press (1975) p. 445.
- k) R. A. Huggins, "Ionically Conducting Solid State Membranes", to be published in Adv. in Electrochem. and Electrochem. Eng., Vol. 10 (1976).

HIGHER ORDER PROCESSES
IN TUNNELING BARRIER DIODES

A. Yariv

ABSTRACT

Frequency Response Limitations due to high order mixing processes in tunnel diode mixers.

In addition to frequency response limitations inherent to quantum mechanical tunneling which are reviewed by M. Tinkham we find a new limitation which is peculiar to the interaction between the quantum mechanical tunneling and the circuit frequency response. If as an example we are interested in the process whereby the frequency $\omega_1 \pm \omega_2 \pm \omega_3$ is generated by mixing the fields $v_1 e^{i\omega_1 t}, v_2 e^{i\omega_2 t}, v_3 e^{i\omega_3 t}$ then the actual fields appearing across the junction will be

$$\frac{v_1}{(1+i\omega_1 t_1)}, \frac{v_2}{(1+i\omega_2 t_2)}, \frac{v_3}{(1+i\omega_3 t_3)}$$

where t_1, t_2, t_3 are effective circuit RC time constants as discussed by Richards which in Javan devices are $\sim 10^{-14}$ sec.

When this fact is incorporated into the basic tunneling expression one finds that the power generated at $\omega_1 \pm \omega_2 \pm \omega_3$ is reduced by

$$\frac{1}{(1+\omega_1^2 t_1^2)^l (1+\omega_2^2 t_2^2)^m (1+\omega_3^2 t_3^2)^n}$$

This would predict a steep penalty for high order processes beyond $w \sim 1/t$.

Considerations of Coupling Efficiency from electromagnetic modes to diode antennas.

Basic electromagnetic considerations applied to the problem of coupling losses radiation into the MOM antennas reveal that in general the received power P_r is related to the incident power P_ℓ or

$$P_r = P_\ell G_a P_\ell \frac{\Omega_\ell}{4\pi} = P_\ell G_a \frac{\lambda^2}{4\pi^2 w_o^2}$$

where G_a is the antenna gain, w_o is the minimum radius (waist) of the focused laser beam and Ω_ℓ is the solid angle of the laser beam. For the ordinary experiments reported, it is estimated that the coupling efficiency varied between 0.1% and 1%. The answer to increased coupling efficiency seems to be in increasing the antenna gain G_a and some means for achieving it will be discussed.

HIGHER ORDER PROCESSES
IN TUNNELING BARRIER DIODES

A. Yariv

The process of mixing infrared and microwave frequencies by means of nonlinear metal-oxide-metal barriers was first demonstrated by Javan and coworkers and is amply documented¹.

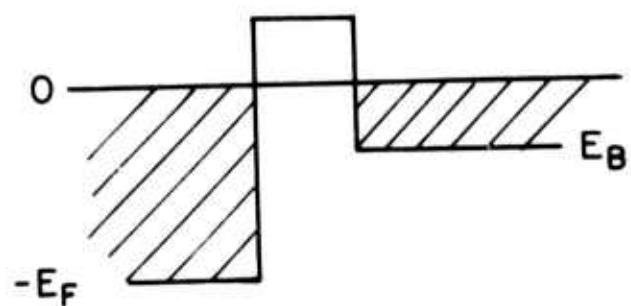
Two intriguing aspects of this work involve: (a) The ability to perform the mixing at ever higher frequencies. (b) The ability to relate the different orders of the observed mixing currents to the corresponding derivatives in the dc V-I characteristics.

We will address both of these problems and consider some important consequences which seem to have gone unnoticed.

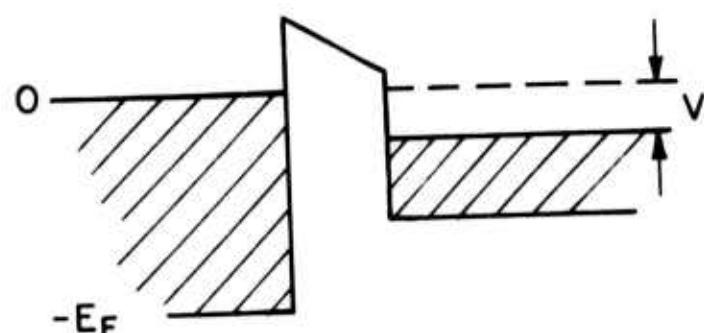
Referring to the model of a tunneling barrier in Fig. 1 the effect of an applied voltage V is to shift the right hand side of the potential energy diagram as in 1(b).

At zero temperature tunneling current is due to electrons with energies between $-V$ and zero (cross hatched in Fig. 1(b)) which find themselves lined up against empty states across the barrier.

The total current is then



1 (a)



1 (b)

Figure 1. The electron potential energy diagram in a metal barrier junction with, 1(a), and without, 1(b), an applied voltage V .

$$I(V) = e \int_{-V}^0 \rho(E)T(E,V)dE \quad (1)$$

where the zero energy reference is taken at the Fermi level on the left side of the barrier. $\rho(E)$ is the density of states and $T(E,V)$ is the tunneling probability per unit time for electron of energy E at the presence of a voltage V . We note that V is the voltage appearing across the physical barrier and not the voltage applied across the external terminals. The importance of this point will become clear below.

We will assume in what follows that the inherent frequency response of the tunneling mechanism is not a limiting factor in the frequency response up to frequencies of 10^{15} - 10^{16} Hz. This is about the number one gets by considering a number of various factors which can contribute to this response. For frequencies below this limit one can write (1) as

$$I(t) = e \int_{-V(t)}^0 \rho(E)T(E,V)dE \quad (2)$$

from which it follows that if

$$V(t) = V_0 + \Delta V(t) \quad (3)$$

$I(t)$ can be obtained as a Taylor expansion

$$I(t) = I(V_0) + \frac{\partial I}{\partial V} \Delta V(t) + \frac{1}{2!} \frac{\partial^2 I}{\partial V^2} (\Delta V)^2 + \dots \frac{1}{n!} \frac{\partial^n I}{\partial V^n} (\Delta V)^n \quad (4)$$

where all the derivatives are evaluated at the bias point

$$V = V_0.$$

To be specific we will consider the three photon process

$$\ell\omega_1 + m\omega_2 - q\omega_3$$

which mixes applied voltages

$$V_1 e^{i\omega_1 t}, V_2 e^{i\omega_2 t}, V_3 e^{i\omega_3 t}$$

to generate a current through the diode at $\ell\omega_1 + m\omega_2 - q\omega_3$, where ℓ, m, q are arbitrary integers[†].

Let us take $\Delta V \equiv a + b + c$ (where $a \equiv V_1 e^{i\omega_1 t}$, $b \equiv V_2 e^{i\omega_2 t}$, $c \equiv V_3^* e^{i\omega_3 t}$) in the n^{th} term of (4).

$$(\Delta V)^n = (a+b+c)^n = \sum_{s=0}^n \sum_{r=0}^s \binom{n}{s} \binom{s}{r} a^{n-s} b^s c^r \quad (5)$$

where () stands for the binomial coefficient. The frequency component $\ell\omega_1 + m\omega_2 - q\omega_3$ will thus result from taking $n = \ell + m + q$, $s = m + q$, $r = q$, so that (5) becomes

$$\begin{aligned} & \left[V_1 e^{i\omega_1 t} + V_2 e^{i\omega_2 t} + V_3^* e^{-i\omega_3 t} \right]^{\ell+m+q} \\ &= \binom{\ell+m+q}{m+q} \binom{m+q}{q} \left(V_1 e^{i\omega_1 t} \right)^\ell \left(V_2 e^{i\omega_2 t} \right)^m \left(V_3^* e^{-i\omega_3 t} \right)^q \\ &= \frac{(\ell+m+q)!}{\ell! m! q!} (V_1)^\ell (V_2)^m (V_3^*)^q e^{i(\ell\omega_1 + m\omega_2 - q\omega_3)t} \quad (6) \end{aligned}$$

[†]We will simplify our algebra by taking ℓ, m, q as positive and allow the frequencies to be positive or negative so that $\ell\omega_1 + m\omega_2 + q(-\omega_3)$. The voltage of a negative frequency term is always to be taken as the complex conjugate.

Substitution of the last expression in the $n = \ell+m+q$ term of (4) gives

$$I_{(\ell+m+q)} = \frac{1}{\ell!m!q!} \frac{\partial^{\ell+m+q} I}{\partial V^{\ell+m+q}} (V_1)^\ell (V_2)^m (V_3)^q e^{i(\ell\omega_1 + m\omega_2 - q\omega_3)t} \quad (7)$$

We emphasize again that ℓ , m and q in (7) are positive integers regardless of whether the corresponding frequencies are added or subtracted.

The last expression brings out a special feature of the high frequency roll-off characteristics of the tunnel diode. Recalling that V_1 , V_2 and V_3 are the amplitudes of the voltages across the physical barrier, we need to relate them to the applied voltages V_{a1} , V_{a2} , V_{a3} .

Consider the equivalent circuit of the antenna and the diode shown in Fig. 2. Here the voltage $V_{ai} = \sqrt{8P_r r_a}$ where P_r is the power available from the antenna at ω_i . The equivalent antenna source impedance is r_a . r_s is the parasitic series resistance of the diode and its connections, R is the diode differential resistance at ω_i at the bias point and C the capacitance of the barrier. The actual voltage across the junction is thus

$$V_i = \frac{V_{ai}}{\left(1 + R_s/R\right) \left(1 + \frac{i\omega CR_s}{R}\right)} \quad (8)$$

where $R_s = r_a + r_s$.

The roll-off (half-power) characteristic frequency is

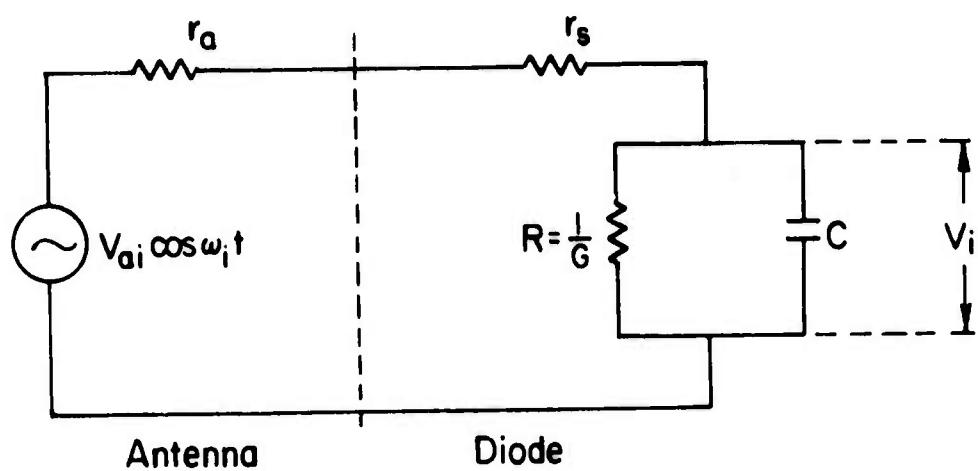


Figure 2. Equivalent circuit of an antenna connected to a diode.

thus

$$\omega_{\frac{1}{2}} = \frac{1 + R_s/R}{CR_s} \quad (9)$$

and is thus $\omega_{\frac{1}{2}} \sim \frac{1}{R'C}$ where R' is the smaller of R_s and R . In practice r_a , the antenna resistance is $50-100\Omega$ so that optimum operation results for $R \approx r_a$ (and $r_s \ll r_a$) so that $\omega_{\frac{1}{2}} \sim \frac{1}{r_a C}$.

In the case of the point contact metal barrier metal (MBM) diodes the minimum capacitance is limited by the minimum spot radius ($\sim 1000\text{\AA}$) and the barrier thickness of $\sim 6\text{\AA}$ to $C \approx 10^{-15}$ farad. Using $R' \approx r_a = 100\Omega$ we obtain $\tau = R'C = 10^{-13}\text{s}$, $\omega_{\frac{1}{2}}/2\pi = 1/(2\pi\tau) \sim 1.5 \times 10^{12}$ Hz for the roll-off frequency.

Returning to the mixing problem - if we express the voltages V_1 , V_2 , V_3 in terms of applied voltages V_{a1} , V_{a2} , V_{a3} we obtain from (7) and (8)

$$I^{(l+m-q)} = \frac{1}{l!m!q!} \frac{\partial^{l+m+q} I}{\partial V^{l+m+q}} \frac{(V_{a1})^l (V_{a2})^m (V_{a3}^*)^q e^{i(l\omega_1+m\omega_2-q\omega_3)t}}{(1+R_s/R)^{l+m+q} (1+i\omega_1\tau)^l (1+i\omega_2\tau)^m (1-i\omega_3\tau)^q} \quad (10)$$

where for simplicity we neglected the possible frequency dependence of $\tau \equiv R'C$ and R_s .

The term $(1 + R_s/R)^{l+m+q}$ indicates the penalty paid when $R \ll R_s$. The second feature is that of the frequency roll-off represented by the factor

$$\eta = \frac{1}{(1+i\omega_1\tau_1)^l (1+i\omega_2\tau_2)^m (1-i\omega_3\tau_3)^q} \quad (11)$$

This shows that the frequency roll-off will be governed by the largest of ω_1 , ω_2 , and ω_3 . Taking, arbitrarily, the highest

frequency as ω_1 , we have that the power at $\ell\omega_1 + m\omega_2 - q\omega_3$ will decrease as

$$|\eta|^2 \propto \left(\frac{1}{\omega_1 \tau_1}\right)^{2\ell} \quad \text{for } \omega_1 > \frac{1}{\tau_1} \quad (12)$$

Since the time constant τ_1 was shown to be approximately 10^{-13} s for typical MBM diodes, we may expect significant roll-off to occur once wavelengths near $10-50\mu\text{m}$ are approached.

The Importance of Sequential Processes

A given frequency ω_4 can be synthesized by a process

$$\omega_4 = \ell\omega_1 + m\omega_2 - q\omega_3$$

involving the direct mixing of the three fields as in (10). Other channels, however, for generating this frequency exist. One such process, as an example, would involve the generation of a signal at $(\ell-2)\omega_1 + m\omega_2 - q\omega_3$, due to a direct $(\ell-2)+m+q$ process and its subsequent beating with a field at $2\omega_1$ due to a second order process. The number of such sequential processes increases rapidly with the order $\ell+m+q$ and the question arises as to how efficient they are in comparison to the direct processes.

Let us become specific and consider the process of generating, as an example, the fifth harmonic of an applied field $Ve^{i\omega t}$. The direct process leads to a current which we can obtain from (10) by taking $\ell=5$, $m=0$, $q=0$

$$I^{(5\omega)} = \frac{1}{5!} \frac{\partial^5 I}{\partial V^5} V^5 e^{i5\omega t} \quad (13)$$

But at the same time there exists a current at 3ω given by

$$I^{(3\omega)} = \frac{1}{3!} \frac{\partial^3 I}{\partial V^3} V^3 e^{i3\omega t} \quad (14)$$

This current causes a voltage

$$V' (3\omega) = \frac{1}{3!} \frac{\partial^3 I}{\partial V^3} V^3 Z^{(3\omega)} e^{i3\omega t} \quad (15)$$

to appear across the barrier where $Z^{(3\omega)}$ is the impedance presented to this current at the diode terminals. In a like manner a voltage

$$V' (2\omega) = \frac{1}{2!} \frac{\partial^2 I}{\partial V^2} V^2 Z^{(2\omega)} e^{i2\omega t} \quad (16)$$

is produced. The voltages $V' (2\omega)$ and $V' (3\omega)$ mix via the term $\frac{1}{2} \frac{\partial^2 I}{\partial V^2} (\Delta V)^2$ in (4) to yield a current

$$I' (5\omega) = \frac{1}{3!} \frac{1}{2!} \frac{\partial^3 I}{\partial V^3} \left[\frac{\partial^2 I}{\partial V^2} \right]^2 V^5 Z^{(3\omega)} Z^{(2\omega)} e^{i5\omega t} \quad (17)$$

The measured current at 5ω is the sum of (13) and (17) as well as of other indirect processes. The ratio of the sequential current $I' (5\omega)$ to the direct one is thus

$$\frac{I' (5\omega)}{I (5\omega)} = \frac{5!}{3! 2!} \frac{\frac{\partial^3 I}{\partial V^3} \left[\frac{\partial^2 I}{\partial V^2} \right]^2 Z^{(3\omega)} Z^{(2\omega)}}{\frac{\partial^5 I}{\partial V^5}} \quad (18)$$

A comparison of I' and I thus requires a knowledge of the various derivatives and of the impedances. To make a numerical

estimate let us assume that the V-I characteristics are exponential

$$I = I_o (e^{SV} - 1) \quad (19)$$

we then have

$$\frac{\partial^n I}{\partial V^n} = I_o S^n e^{SV} \quad (20)$$

and

$$\frac{I_{(5\omega)}}{I_{(5\omega)}} = 10 I_o^2 S^2 e^{2SV} z_{(3\omega)} z_{(2\omega)} \quad (21)$$

if the impedances at the various ω are the same and are given by

$$z(\omega) = \left(\frac{\partial I}{\partial V} \right)^{-1} = \left(I_o S e^{SV} \right)^{-1} \quad (22)$$

It follows from (21) and (22) that

$$\frac{I_{(5\omega)}}{I_{(5\omega)}} = 0$$

i.e., the sequential process could be larger than the direct process provided the impedances are comparable and subject to the exponential characteristic assumption. The point we wish to make, however, is mainly that it is dangerous to relate the observed mixing currents to individual terms in V-I series expansion and that in reality we may be observing the net result of a number of processes. Some added information is necessary to tie down a given frequency to the diode characteristics.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

Reference

1. A. Javan in "Modern Methods in Precision Spectroscopy," Proceedings of the Esfahan Symposium, Edited by M. Feld and A. Javan, J. Wiley, Interscience, 1973.

SUMMARY OF MEETING ON "OPTICAL ELECTRONICS"

N. Bloembergen

ABSTRACT

A two-day meeting was held on the current state-of-the-art and the potential of infrared barrier-diodes. The nonlinear properties, response time and detection sensitivity of submillimeter and infrared Schottky-barrier, point-contact and metal-oxide-metal devices were reviewed and compared with theoretical predictions. The results are summarized in a number of conclusions and the following recommendation for further research and development: Current activities should receive continued support. Additional support should be provided to encourage groups with broader expertise in engineering and material science aspects to participate. Micro-engineering of barriers and optimization of devices toward specific applications should be emphasized.

SUMMARY OF MEETING ON "OPTICAL ELECTRONICS"

N. Bloembergen

At the suggestion of Dr. C. M. Stickley of ARPA a meeting was convened to review research programs in "Optical Electronics". This rather broad term is used in the present context to indicate the extension of lumped circuit element techniques, well known in the radio- and microwave frequency region, to the infrared and visible region of the electromagnetic spectrum. In practice, the discussion was limited to barrier diode devices. Such devices have been used quite successfully as frequency mixers in the infrared to permit absolute frequency measurements of infrared laser lines. They can be shaped to physical dimensions comparable to, or smaller than, infrared wave lengths. Their potential as infrared detectors, as fast logic elements, and perhaps as infrared oscillators should be assessed.

The devices can take on various physical forms, but basically always contain two conducting materials, separated by a narrow insulating barrier. A metal may be contacted with a heavily doped semiconductor (GaAs or Si). A Schottky barrier will be formed by depletion of the carriers in the space charge layer near the semiconductor-metal interface.

A point contact diode is formed by pressing a pointed

wire tip (for example tungsten) against an oxidized metal surface (nickel). The contact area may have a diameter smaller than 10^{-5} cm, while the oxide layer may be only 10^{-7} cm (10 angstroms) thick. Such devices are relatively easy to make, by trial and error. They are, however, extremely fragile, unstable and rather irreproducible. With micro-electronic photo-resist and high vacuum evaporation techniques, it is possible to make more reproducible and stable configurations, although as yet not of quite as small dimensions as the point contact. One can, for example, deposit a fine strip of aluminum, oxidize it with a layer of controlled thickness, and evaporate another thin aluminum strip at right angles. Such techniques also permit the construction of diode arrays. The insulating layer need not be an oxide (MOM structure). It could be an insulating organic molecular layer (MIM structure) or even a semiconductor layer (MSM structure). Such structures would, in principle, permit better controlled variations of the barrier potential. The generic term, metal-barrier-metal (MBM) structures, appears more appropriate.

Furthermore, the metal on one or on both sides of the barrier may be superconducting. In the latter case both normal one-electron tunneling and superconductive pair-tunneling may occur, leading to Josephson effects. The theory of electron tunneling through MBM and Josephson functions is well developed at microwave frequencies. So is the theory of electron emission over a Schottky barrier. The nonlinear characteristics

of the current-voltage response are well known.

The extension to infrared and optical frequencies involves a consideration of the following factors:

1. The impedance matching of the incident infrared radiation field to the diode structure. One of the conductors should be shaped in the form of an antenna, while the other conductor can serve as a ground plane.

2. The high-frequency roll-off characteristics of the equivalent lumped circuit, consisting of a spreading resistance R_s in series with a parallel RC combination, incorporating the resistance and capacitance of the barrier proper.

3. The quantum effects that occur when the photon energy $\hbar\omega$ becomes comparable to, or larger than, the barrier potential. A transition takes place from the tunneling characteristic to photo-emissive conduction. In the case of the Schottky barrier the current-voltage characteristic changes, when $\hbar\omega$ becomes larger than kT .

These three points have been addressed in more detail in technical papers prepared by the MRC members, P. L. Richards, M. Tinkham and A. Yariv. Their reports may be found elsewhere in the present volume.

This broadly sketched picture partly resulted from and was partly reinforced by the presentations and discussions at the meeting, during which MRC members met with approximately twenty participants from interested government, academic and industrial organizations. The formal program was composed

as follows:

Opening Remarks, N. Bloembergen

ARPA Interest in Optical Electronics, C. M. Stickley

"State of the Art of Optical Electronics", A. Javan

"Fundamental Limits to the Sensitivity of High-Frequency Josephson Effect Receivers", P. L. Richards

"Josephson Effect Memories", N. Welker

"Frequency Measurements with Tunnel Diodes at the National Bureau of Standards", K. Evenson

"IR Detection and Mixing with Heavily Doped Barriers", A. Van der Ziel

"Quantum Theory of Mixing", J. Tucker

"Schottky Barrier Mixing Experiments", A. Silver

"Tunnel Diodes and Guided Light Waves", S. Wang

"Nonlinear Characteristics of Tunnel Diodes", K. Gustafson

"Experimental Configurations of Point Contact Diodes", S. E. Schwartz

General Round Table Discussion "Are Optical Diodes Ripe for Technological Development and Exploitation?"

Dr. Stickley (ARPA) pointed out that one of the most important areas of ARPA interest in these devices is the potential use as room-temperature detectors for $10.6\mu\text{m}$; or more generally $8-14\mu\text{m}$, radiation. Of particular importance is the possibility of forming images by means of diode arrays.

Professor Javan (MIT) gave an account of the history of IR frequency mixers. He pointed out that MOM point contacts with a diameter of $0.1\mu\text{m}$ and a barrier thickness of

about 10 angstroms have an RC time constant, $T \sim 10^{-14}$ sec. This permits frequency mixing from microwaves and millimeter waves with various infrared laser lines. The chain of frequency mixing has been extended to the near infrared. He also showed the response of a barrier diode to five visible frequencies from an argon ion laser. The barrier height derived from the photo conductive response at visible frequencies corresponded well with the barrier height derived from tunneling characteristics at microwave frequencies. The complex dependence on frequency and dc bias voltage appears well understood. Javan also presented some ideas on how phase and amplitude information of infrared radiation may be collected by an array of dipoles illuminated with a coherent local oscillator IR laser. The heterodyne mixing would transfer the holographic information to microwave channels.

Professor Richards (MRC) reviewed the fundamental limits of thermal and quantum noise of Josephson effect detectors and compared these to other submillimeter and IR detectors. He gave an analysis which suggested that present MBM diodes have NEP sensitivities far below the theoretical limits because of poor coupling between the incident radiation with the device and because of excess noise.

Dr. N. Welker (NSA) reviewed work on Josephson junction memory devices, carried out mostly at the IBM research laboratories. These devices have a negative resistance characteristic and combine low power consumption with a fast

switching rate. A power consumption of one microwatt per element with a switching rate of 10^8 was quoted.

Dr. K. Evenson (NBS, Boulder) described infrared mixing techniques at $2\mu\text{m}$ wave length. This extension of point contact mixing techniques makes precise frequency measurements possible throughout most of the infrared region. The response characteristics of different existing detectors, were compared and experimental data are tabulated in Table 1. It should be emphasized that no corrections were made for insertion losses, which in several cases are quite large. The dynamic range and burn-out power are additional important features not included in the table.

There are fundamental reasons to expect that very non-linear devices such as super-Schottky and Josephson junctions will display a roll-off at lower frequencies. There are also apparent trade-offs between response time and sensitivity, expressed in terms of noise-equivalent power (NEP).

Professor A. Van der Ziel (University of Minnesota) presented a detailed theory of infrared mixing in heavily doped Schottky barriers. The frequency response of these devices could in principle be extended to the infrared.

Dr. A. Silver (Aerospace Corp.) described the successful operation of GaAs Schottky barrier devices at submillimeter wave lengths. Fabrication of diodes with Au contact points of $0.5\mu\text{m}$ diameter with a spacing of $3-6\mu\text{m}$ between diodes has been achieved. The spreading resistance is the limiting

Table 1. Table of Detector Characteristics

Detector Type	Wave length limit in μm	NEP watt/Hz	Upper mixing frequency limit (10^{12} Hz)	Characteristic response time (10^{-12} sec.)	Observed harmonic Mixing Order
Silicon Point Contact	300	10^{-11}	2.5	0.4	>200
Schottky Barrier (GaAs)	10	10^{-8}	2.5	>0.4	33
Josephson Junction	10	10^{-14}	2.8	0.3	825
MIM Point Contact	2	10^{-8}		0.01	15
MIM Evaporated Layer	10	10^{-7}	0.3	>1	
Photoconductive (doped-Ge)	$10^{-0.2}$	10^{-15}		>1000	

factor. To obtain a sufficient concentration of carriers ($3 \times 10^{18}/\text{cc}$), p-type GaAs must be used. A promising material to get a high carrier concentration with a higher mobility is InSb. Superconducting Pb-GaAs junctions have been used and the characteristics of such super-Schottky barriers have been measured.

Dr. J. Tucker (Aerospace Corp.) described the theory of high-frequency operation of Schottky barriers. Quantum effects become important when the photon energy becomes large compared to kT . In the high frequency limit a photo-emissive type of response is obtained. This question is also discussed in more detail in a MRC paper by M. Tinkham.

Professor S. Wang (University of California, Berkeley) addressed the question how linear arrays of diodes may be used to load light wave guides. The integration of diode devices with other elements of integrated optics is intriguing.

Professor K. Gustafson (University of California, Berkeley) showed detailed experimental response curves for MBM devices. The nonlinear characteristics observed are in good agreement with the theory of tunneling. The dependence on the angle of incidence and on the polarization of the incident radiation is in agreement with the application of long thin-antenna response theory to the contacting wire. Negative resistance characteristics appeared to be evident at low frequencies, but it is quite uncertain whether the negative resistance domain will persist at microwave frequencies.

Professor S. E. Schwartz (University of California, Berkeley) considered the negative resistance characteristics that might result from space charge interactions in the barrier of the diodes. He proposed to utilize this effect for the construction of low power local oscillators. He also discussed the properties of MBMBM sandwich structures, which may lead to the construction of triodes and other devices.

In general discussion at the end of the meeting the need for constructing reproducible barriers was stressed. This task becomes more difficult as the dimensions get smaller. The effects of pressure, temperature and impurities on the barrier operation have not been investigated systematically.

The results of the meeting may be summarized in the following conclusions and recommendations.

1. The basic physical mechanisms operating in IR barrier diodes are known. There is no fundamental obstacle to constructing diodes with response times of 10^{-14} sec.
2. Frequency mixing down to $2\mu\text{m}$ wavelength has been achieved, and may possibly be extended to the visible region, although the frequency roll-off characteristic for high order mixing processes are fast and the quantum effects leading to photo-emission will wash out the nonlinear tunneling characteristics.
3. The present sensitivity of fast diode detectors is many orders of magnitude below the theoretical limit. Considerable improvement may be obtained by analysis of the

antenna impedance matching characteristics and the reduction of input coupling losses.

4. Further development in the fabrication and design of reproducible barrier contacts of small dimensions and with small spreading resistance is needed. Barrier layers of semi-conducting materials and organic molecules should be considered. Further optimization of existing Schottky diode fabrication technology for IR applications should be fruitful. Deposition of barriers under ultra-clean high vacuum conditions and with molecular beam techniques should be considered.

5. The power handling capabilities of these small devices is obviously severely limited. The effects of heating by local oscillator mixing power should be analyzed and basic limitations imposed by burn-out should be established.

6. The design and development of a prototype fast room-temperature MBM detector for $10.6\mu\text{m}$ radiation may be undertaken immediately.

7. Extensive work on arrays and incorporation of diodes with integrated optics circuits should be paced according to the performance achieved in the MBM diode IR detector.

8. Negative resistance characteristics for superconducting junctions are firmly established, but further confirmation is required for their existence in other barrier devices. Their potential as fast logic devices cannot be accurately assessed at the present time. The use of these devices as IR local oscillators appears limited.

9. The field of fast IR diodes holds sufficient promise that further basic research should be encouraged. Design studies for large scale arrays and their applications should be undertaken, although the field is not ripe for the construction of large scale arrays.

10. Current activities in this field should receive continued support. Additional support should be provided to encourage groups with broader expertise in engineering and material science aspects to participate. Micro-engineering of barriers and optimization of devices toward specific applications should receive emphasis.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

FUNDAMENTAL CONSIDERATIONS
LIMITING TUNNEL DIODE MIXERS AND DETECTORS

M. Tinkham

ABSTRACT

Fundamental physical principles limit the degree to which one can simultaneously achieve high speed (or maximum useable frequency), high sensitivity, and high temperature of operation of tunnel diode detectors. In the context of a voltage-controlled mixer or square-law detector, a useful figure of merit is the logarithmic derivative $S = \frac{d}{dV} \ln \frac{dI}{dV} = (d^2I/dV^2)/(dI/dV)$, which has the dimensions of volts^{-1} . In a classical Schottky diode, for example, $S = e/kT$, so that sensitivity apparently can be increased without limit by lowering the temperature. In fact, this increase reaches a limit when thermal electron currents drop below the tunneling current. Moreover, as shown by Tucker, the improvement which can be obtained in this way is also limited by quantum effects as soon as $kT \sim \hbar\omega$; specifically, S approaches $2e/\hbar\omega$ for $(kT/\hbar\omega) \ll 1$, corresponding to a single extra electron of current per photon absorbed. None the less, further cooling below $T = \hbar\omega/k$ may be needed to reduce background current. By contrast, in the MOM tunnel diode there is no sharp feature in the density of electronic states on the energy scale of kT (such as the semiconductor or superconductor gap edge), and so S

must be of the order of the inverse of the barrier height V_b , in volts. Thus, $S \approx 1 \text{ volt}^{-1}$, compared to $\approx 40 \text{ volt}^{-1}$ for a room temperature Schottky diode, or $\approx 12,000 \text{ volt}^{-1}$ for a super-Schottky at 1°K . This comparison explains the poor sensitivity of the MOM diode detector or mixer; but it also provides the basis for the MOM's remarkable frequency response, which is flat out to photon energies $\hbar\omega \approx eV_b \approx 1 \text{ eV}$. In fact, we argue that there is a quantum limit on the nonlinearity-bandwidth product of the form $S_0 \Delta\omega < 2e/\hbar$, where S_0 is the low-frequency value of S and $\Delta\omega$ is the bandwidth. That is, it is not possible to obtain the fast response of a MOM diode except at the expense of low sensitivity, even if one allowed operation at low temperatures. For example, it appears that such sharp features as the zero-bias anomaly in the I-V curve can not be effective in giving a sensitive and fast response. However, it may be possible to make considerable improvement in existing diodes by optimizing the barrier design tradeoffs for a given speed, frequency, and temperature of operation. Specifically, it appears that there is no fundamental reason to prevent development of a Schottky tunnel diode using heavily doped material (such that the plasma frequency is comparable to the frequency to be detected), operating at room temperature, which will require local oscillator power of no more than 10^{-4} Watts to approach the quantum limit for heterodyne detection sensitivity at $10\mu\text{m}$. Such a device would clearly be of great practical value for heterodyne systems using CO_2 laser radiation.

FUNDAMENTAL CONSIDERATIONS LIMITING
BARRIER DIODE MIXERS AND DETECTORS

M. Tinkham

I. Introduction

Any circuit element with a non-linear I-V characteristic can be used for frequency mixing and detection. There is naturally a premium on those which are fast, efficient, low-noise, and can operate with minimum refrigeration. Unfortunately these desirable features are mutually incompatible, and compromises must be made. Although these tradeoffs are presumably similar in all forms of detectors, in this paper we shall restrict our attention to barrier diodes of various sorts. This restriction is reasonable since MBM (metal-barrier-metal), Schottky (metal-depletion layer barrier-semiconductor), and super-Schottky (superconductor-depletion layer barrier-semiconductor) diodes are some of the most promising high performance devices. Moreover, the barrier diode configuration lends itself to analysis in fundamental terms.

In a barrier diode, two regions of highly conducting material are separated by a thin barrier through which electrons can not flow by ordinary conduction at the Fermi level. Instead, they either must cross the barrier by the quantum mechanical tunnel process or else go over the top by thermal activation.

Roughly speaking, for barrier heights well above kT and for small applied voltages V , the net tunnel current density is

$$J_{\text{tun}} = (ne^2/mv_F)Ve^{-2(2meV_b)^{1/2}d/\hbar} \quad (1a)$$

while the net thermal current over the barrier is

$$J_{\text{th}} = (ne^2/mv_F)Ve^{-eV_b/kT} \quad (1b)$$

where V_b is the barrier height in volts, d is the barrier thickness, T is the temperature, and e , m , \hbar , and k are the usual fundamental constants. The common coefficient is

$$ne^2/mv_F = \sigma/\ell = (\rho\ell)^{-1} = \omega_p^2/4\pi v_F \approx 0.8 \times 10^{-4} n^{2/3} \text{ohm}^{-1} \text{cm}^{-2} \quad (1c)$$

n being the carrier concentration per cm^3 . Note that both (1a) and (1b) reduce to $J = \sigma E$ if V_b is set to zero and V is taken to be the voltage per mean free path. Comparing (1a) and (1b), we see that the tunnel current dominates when

$$(eV_b)^{1/2} \left(\frac{\hbar^2}{8md^2} \right)^{1/2} > kT \quad (1d)$$

Even at room temperature, where $kT = (1/40)\text{eV}$, tunneling will dominate for a 10\AA barrier if $V_b \gtrsim 0.08\text{V}$, and for a 50\AA barrier if $V_b \gtrsim 2\text{V}$. At lower temperatures, thermal activation over the barrier is even less of a factor. Accordingly, in many situations of interest it will be appropriate to treat the current as arising from a tunneling process.

For the important case of a Schottky diode where the barrier results from a depletion layer, the maximum barrier height and thickness are related by $V_b = 2\pi N_d e d^2$, where N_d is the number of donors per unit volume. In this case, the tunneling and thermal currents both are independent of d and fall exponentially with V_b [as $\exp(-m/\pi\hbar^2 N_d)^{1/2} V_b$ and $\exp(-e/kT) V_b$, respectively], so that the crossover temperature depends only on N_d , namely

$$T_{\text{crossover}} = \frac{e\hbar}{k} \left(\frac{\pi}{m}\right)^{1/2} N_d^{1/2} = \hbar\omega_p/2k \approx 2 \times 10^{-7} N_d^{1/2} \quad (1e)$$

(We have used the free electron mass in the numerical coefficient.) For example, in the heavily doped GaAs quoted by the Aerospace group, $N_d = 5 \times 10^{18} \text{ cm}^{-3}$, and this crossover temperature would be $\sim 475^\circ\text{K}$. Thus, in these diodes tunneling is important at room temperature, and completely dominant at reduced temperatures. More generally, this calculation indicates that tunneling will dominate over thermal activation in Schottky diodes when the thermal energy kT is less than the plasma energy $\hbar\omega_p$ of the semiconductor.

In the case of the MBM diode, it is an excellent approximation (for infrared and lower frequencies) to neglect potential differences within the metal electrodes, and consider the entire applied voltage V to appear across the tunnel barrier. In this case it is clear that the appropriate analysis is exactly the usual voltage-controlled one, described by a non-linear function $I(V)$. In the case of the Schottky diodes, one must

take account of the voltage drop in the spreading resistance or the semiconductor side. This is of crucial importance, especially at high frequencies, but it can be treated separately by circuit theory as a linear resistor in series with the voltage-controlled non-linear tunnel resistance.

II. Characterization of Performance

For simplicity, we shall restrict our attention to square-law detection and to heterodyne mixing of two nearby frequencies, leaving aside higher-order mixing processes. With this restriction, we can for low frequencies and low amplitudes characterize the diode adequately by

$$I = b_1 V + b_2 V^2 \quad (2)$$

Then if the applied voltage is

$$V = V_0 + V_1 \cos \omega t \quad (3)$$

the current is

$$I(t) = b_1 V_0 + b_2 \left(V_0^2 + \frac{V_1^2}{2} \right) + b_1 V_1 \cos \omega t + \frac{1}{2} b_2 V_1^2 \cos 2\omega t \quad (4)$$

Thus, the detected dc (short-circuit) current response is

$$\delta \bar{I} = b_2 \frac{V_1^2}{2} = b_2 \bar{V_1^2} = \frac{1}{2} \frac{d^2 I}{dV^2} \Big|_{dc} \bar{V_1^2} \quad (5)$$

The corresponding open-circuit voltage is

$$\delta \bar{V} = R_D \delta \bar{I} \quad (6)$$

where the differential resistance at the operating point R_D is given by

$$R_D = (d\bar{I}/dV_0)^{-1} = (b_1 + 2b_2 V_0)^{-1} \quad (7)$$

The available detected signal power to a matched load R_D is then

$$P_{\text{avail}} = \frac{(\delta\bar{V})(\delta\bar{I})}{4} = \frac{1}{4}(\delta\bar{I})^2 R_D = \frac{b_2^2}{4} \overline{(V_1^2)^2} R_D \quad (8)$$

The corresponding power absorbed from the incident signal is

$$P_{\text{in}} = \overline{I(t)V_1 \cos\omega t} = \frac{1}{2} b_1 V_1^2 = b_1 \overline{V_1^2} \quad (9)$$

Thus, one can define a level-dependent conversion efficiency by

$$\eta \equiv \frac{P_{\text{avail}}}{P_{\text{in}}} = \frac{1}{4} \frac{b_2^2}{b_1} \overline{V_1^2} R_D \quad (10)$$

This analysis can be applied to mixing of two signals $V_1 \cos\omega_1 t$ and $V_2 \cos\omega_2 t$, ($V_1 \ll V_2$), by recognizing that the sum can be viewed as having an instantaneous amplitude varying between $(V_2 + V_1)$ and $(V_2 - V_1)$ at the beat frequency $(\omega_1 - \omega_2)$. The corresponding beat (or i.f.) frequency current has a peak value

$$I_{\text{if}} = b_2 \overline{\delta V_2^2} = b_2 V_1 V_2 \quad (11)$$

Thus, the available power is

$$P_{\text{avail}} = \frac{1}{8} I_{\text{if}}^2 R_D = \frac{1}{8} b_2^2 V_1^2 V_2^2 R_D$$

while the input power is again $P_{in} = \frac{1}{2}b_1V_1^2 = b_1\overline{V_1^2}$, while the local oscillator power is

$$P_{LO} = \frac{1}{2}b_1V_2^2 = b_1\overline{V_2^2} \quad (12)$$

Thus

$$P_{avail} = \frac{1}{2}\left(\frac{b_2}{b_1}\right)^2 P_{LO} P_{in} R_D \quad (13)$$

so that the conversion efficiency is

$$\eta \equiv \frac{P_{avail}}{P_{in}} = \frac{1}{2}\left(\frac{b_2}{b_1}\right)^2 P_{LO} R_D \quad (14)$$

which depends only on P_{LO} , not on the signal power P_{in} .

Superficially, (14) suggests that actual conversion gain could be obtained simply by increasing P_{LO} sufficiently. In practice, P_{LO} is usually limited by heating of the device, or by excess noise introduced by the oscillator. Even if these practical problems did not intervene, a theoretical limit to P_{LO} in (14) follows from the fact that (b_1/b_2) will typically give the order of magnitude of the voltage V_m at which the expansion (2) fails, since successive terms are then equal. Thus, (14) can not be expected to hold beyond $P_{LO} \approx b_1 V_m^2 = b_1 (b_1/b_2)^2$, at which point $\eta = \frac{1}{2}R_D b_1 \approx \frac{1}{2}$. Actually, $\eta = \frac{1}{2}$ is the theoretical limit for an ideal resistive mixer, corresponding to loss of the energy in the sum frequency in the mixing process.

The above analysis suggests the usefulness of defining a normalized curvature parameter S by

$$S = \frac{2b_2}{b_1} = \frac{d^2I/dV^2}{dI/dV} \quad (15)$$

This has the dimensions of Volt⁻¹, and is a measure of the voltage scale over which the curvature of the I-V characteristic changes the local slope. When dealing with more general I(V) curves, in which d^2I/dV^2 and dI/dV are functions of V, S is understood to be evaluated at the operating point V_0 , under high frequency conditions. In terms of this parameter, the heterodyne conversion gain can be written as

$$\eta = \frac{1}{8} P_{LO} R_D S^2 \quad (16)$$

It is useful to note that this implies that the local oscillator power needed to approach the limit $\eta = \frac{1}{2}$ is $P_{LO} = 4/R_D S^2$, which decreases rapidly with increasing S.

A convenient way of treating the square-law and heterodyne detectors on a common basis is in terms of the minimum detectable power P_{min} for unit signal/noise ratio. That is, we equate the available detector output power ηP_{min} to the thermal noise power kTB in bandwidth B. Using (16), we find

$$P_{min,het.} = \frac{8kTB}{P_{LO} R_D S^2} = \frac{\overline{2i_n^2}}{P_{LO} S^2} \quad (17)$$

where $\overline{i_n^2}$ is the mean square Johnson noise current in R_D in bandwidth B. Of course $\overline{i_n^2}$ should be the total noise current if extra noise (such as shot noise) is present. For comparison,

using (10) we find

$$P_{\min., \text{sq. law}} = 4 \left(\frac{kTB}{R_D S^2} \right)^{\frac{1}{2}} = \frac{2}{S} (i_n)_{\text{rms}} \quad (18)$$

From (17) and (18) it is evident that sensitivity is increased by increasing $R_D S^2$ if the noise is thermal and by increasing S^2 if it is not. In fact, for reasonable ease in coupling, $1/b_1 \approx R_D$ must be within an order of magnitude of $Z_0 = 377\Omega$, the impedance of free space. Thus, in either case, the key controllable parameter is S . Sensitivity increases as S^2 while the local oscillator power required for $\eta \approx \frac{1}{2}$ falls as $1/S^2$. Evidently, it is advantageous to seek a large value for S .

III. Factors Governing S in the Low-Frequency Regime

Semiconductor Diodes:

The classical I-V characteristic of a semiconductor diode based on thermally-excited electrons is

$$I = I_0 (e^{eV/kT} - 1) \quad (19)$$

For this characteristic,

$$S_{\text{thermal}} = \frac{d^2 I / dV^2}{dI / dV} = \frac{e}{kT} \quad (20)$$

independent of I or V . This suggests considering the generalization of (19) to

$$I = I_0 (e^{SV} - 1) \quad (21)$$

in which S can be imagined to have values other than e/kT .

Considering first the thermal case, $S = 40 \text{ volt}^{-1}$ at room temperature, and $S \approx 12,000 \text{ V}^{-1}$ at 1°K .

From these numbers, it appears on the face of it that enormous gains in sensitivity can be obtained by taking a Schottky diode to low temperatures. But it is not so simple, because on cooling, the thermal currents, on which (19) is based, are frozen out (since $I_o \sim e^{-eV_b/kT}$), and tunneling currents dominate. As discussed in the Introduction, for the rather heavy doping level $N_d = 5 \times 10^{18}$ used in the Super-Schottky diodes of the Aerospace group, tunneling is expected to be somewhat more important than thermally-activated current even at room temperature. Hence, it will dominate completely at low temperatures. As indicated in the Introduction, the relation between depletion layer thickness and potential difference causes the tunnel current to be dominated by a factor $\exp(-m/\pi\hbar^2N_d)^{1/2}(V_b-V)$. This will correspond approximately to an S value

$$S_{\text{Schottky-tunnel}} = \left(\frac{m}{\pi\hbar^2N_d} \right)^{1/2} = 5.4 \times 10^{10} N_d^{-1/2} = \frac{e}{kT_{\text{crossover}}} = \frac{2e}{\hbar\omega_p} \quad (22)$$

where $T_{\text{crossover}}$ is as defined in (1e), and ω_p is the plasma frequency. Again, in the numerical factor we have used the free mass. This expression gives $S = 24 \text{ V}^{-1}$ for $N_d = 5 \times 10^{18}$, which is closer to the experimental value 19 V^{-1} than is the room temperature thermal value 40 V^{-1} . One might expect the

observed value of S to be the thermal value e/kT well above the crossover temperature (i.e.), rising to $\sim e/kT_{\text{crossover}} = 2e/\hbar\omega_p$ as a limiting value for $T \ll T_{\text{crossover}}$.

Given the above situation, one might try to raise S by lowering N_d so that ω_p and $T_{\text{crossover}}$ will be reduced. But lowering N_d will increase the spreading resistance as well as the tunnel resistance, both of which will lower the efficiency of the overall circuit, as limited by junction capacitance. Moreover, it seems reasonable that ω_p forms a high frequency limit for the normal operation of a Schottky tunnel diode, since the high S value results from changes in the thickness of the depletion layer in response to the applied voltage. Thus, lowering N_d doesn't seem to be a useful direction for high frequency operation. On the contrary, very highly doped semiconductors seem preferable, to approach M-B-M conditions.

Super-Schottky Diodes:

The "Super-Schottky" diode of the Aerospace group provides a means of surmounting these problems and achieving the thermal $S = e/kT$ even at 1°K , where it is very large. In the super-Schottky diode, the electrons tunnel through the depletion layer into a superconducting metal, rather than a normal one. Because of the energy gap of width $2\Delta \approx 3.5kT_c$ at the Fermi level of a superconductor, no tunnel current can flow at $T = 0$ until $eV = \Delta$. The density of states in the superconductor is singular at the edge of the gap, varying from $N_S(E) = 0$ for $|E| < \Delta$ to

$$N_S(E) = N_n \frac{E}{(E^2 - \Delta^2)^{\frac{1}{2}}} \quad (23)$$

for $|E| > \Delta$, where N_n is the density of states at the Fermi level in the normal state, and E is the energy relative to E_F .

Hence, the tunneling current at $T = 0$, which varies as

$$\int_{\Delta}^{eV} N_S(E) dE, \text{ rises as}$$

$$I = \frac{1}{R} [V^2 - (\Delta/e)^2]^{\frac{1}{2}} \quad (24)$$

which has infinite initial slope at $eV = \Delta$, and smoothly approaches the normal state tunneling current when $eV \gtrsim 2\Delta$.

This perfectly sharp "corner" in the (ideal) I-V curve is rounded at finite temperatures by exponential tails of Fermi functions, so that the normalized curvature becomes essentially $S = e/kT$ for $V \lesssim \Delta/e$. For higher voltages, the I-V curve becomes essentially linear, with only the curvature of the normal Schottky tunnel junction, as estimated in (22); this is negligible on the millivolt scale relevant to the superconducting energy gap and to kT at 1°K . All this behavior is demonstrated in the published I-V curves of the super-Schottky diodes.

Clearly this is a very promising, high-sensitivity device, which succeeds in realizing the theoretical $S = 12,000\text{V}^{-1}$ for 1°K operation. Moreover, the normal impedance level R can be easily controlled to fall in the desirable 100Ω range. The essential basis for the operation of this diode is the existence of an infinite discontinuity in $N(E)$ which causes a "corner" in the $I(V)$ curve at $T = 0$, so that only kT is

broadening the I-V curve. This suggests the possibility of increasing $S = e/kT$ indefinitely by lowering T . Of course, in a real diode, Δ is not perfectly sharp (due to gap anisotropy and material inhomogeneity), so even at $T = 0$ there would be some slight residual rounding. Moreover, as we shall see in Section IV, the quantum energy $\hbar\omega$ essentially provides a minimum width, so that even in an ideal diode at $T = 0$, S can not exceed $2e/\hbar\omega$. Thus, for a given frequency ω , reducing kT below $\hbar\omega$ will not continue to improve S in the same way.

M-B-M Diodes:

Finally, let us turn to the metal-barrier-metal tunnel diodes. By this, we mean a diode in which the barrier layer is chemically different from the metals on either side. Thus, unlike in the Schottky, in this case the barrier width d is essentially independent of the voltage. To a first approximation, the tunnel current is linear in the applied voltage, and [using (1a)] the resistance is roughly

$$R = \frac{\rho l}{A} \exp\left[\frac{2}{\hbar}(2m)^{\frac{1}{2}}(eV_b)^{\frac{1}{2}}d\right] \quad (25)$$

where $\rho l = mv_F/ne^2 \approx 1.4 \times 10^{-6} n^{-2/3}$ (e.s.u.) $= 1.3 \times 10^4 n^{-2/3} \text{ ohm-cm}^2$ $\approx 10^{-11} \text{ ohm-cm}^2$ for a typical metal or 10^{-9} ohm-cm^2 for a degenerate semiconductor, and A is the junction area.

If the barrier and electrodes are completely symmetrical, $I(V)$ must be an odd function, and the leading terms will be

$$I = \frac{1}{R} V \left[1 + \alpha \left(\frac{V}{V_b} \right)^2 + \dots \right]$$

where α will be of order 1, but depend on the details of the barrier. Thus, to get a second-order mixing coefficient, a bias V_o is required, where V_o would be expected to be limited to $\sim (V_b/2)$ by the higher nonlinearities, heating, etc. With such an external bias, or with a "built-in" bias of similar size from an asymmetric junction, the normalized curvature parameter S might be estimated as $\alpha V_o / V_b^2$ or

$$S_{M-B-M} \approx (1/V_b)$$

For typical barrier heights of order 2 volts, $S_{M-B-M} \approx 0.5V^{-1}$, two orders of magnitude below even a room-temperature Schottky diode. The reason that a Schottky diode with a similar value of $V_b \sim 1V$ can have $S_{tunnel} \sim 20V^{-1}$ is that in a Schottky the depletion layer barrier has a thickness which depends in first order on V , whereas in an M-B-M diode d is constant. Thus, the efficiency of an M-B-M diode mixer will typically be four orders of magnitude below that of a room-temperature Schottky, or even one with tunneling playing an important role.

However, the M-B-M diode has the advantage of being useable up to extremely high frequencies, because the low series resistance of the metal electrodes reduces the problem of coupling efficiently into the capacitively-shunted non-linear junction element. Contrary to some opinions, the

"intrinsic" high speed of the tunnel effect does not seem to be the critical factor, since heavily-doped Schottky diodes also operate in a tunneling mode even at room temperature. However, the response speed of a tunnel Schottky will presumably be limited by the plasma frequency of the semiconductor, which should govern the speed of changes in the depletion layer. This plasma frequency $\omega_p = (4\pi ne^2/m)^{1/2}$ would lie around 0.1eV ($\lambda \approx 10\mu\text{m}$) for a degenerate semiconductor, but around 10eV (1000 \AA) for a metal.

These considerations already provide an example of the sort of intrinsic limitations we have in mind. According to (22), $S \approx 2e/\hbar\omega_p$, and now we argue that ω_p sets a sort of high-frequency limit on the operation of the Schottky tunnel diode. This leads to a limit on the nonlinearity-bandwidth product

$$S \Delta\omega \lesssim S \omega_p \approx 2e/\hbar \quad (26)$$

Although we arrived at this result from consideration of a Schottky tunnel diode, it also appears to hold at least roughly for an oxide M-B-M diode, in which $S \sim 1/V_b$ and where photo-assisted processes come in to reduce S for $\hbar\omega \gtrsim eV_b$.

Another consideration leading to a high frequency roll-off is the RC time constant associated with the shunt capacitance of the diode, which will tend to bypass the high-frequency current around the non-linear resistance. If the diode were truly voltage controlled, such shunt effects would be irrelevant, but in practice the diode always has series resistance

from the radiation resistance of the antenna driving it, as well as the resistance of the metal leads or, worse yet, the spreading resistance in the semiconductor side of a Schottky diode. Accordingly, it is of interest to compute the RC time constant of the junction itself. Since $C = \epsilon A / 4\pi d$, our estimate (25) of R leads to

$$\tau = RC = \frac{\epsilon \rho \ell}{4\pi d} e^{\alpha V_b \frac{\lambda}{2} d} \approx \frac{2 \times 10^{-15} \text{ sec}}{d} e^{1.1 V_b \frac{\lambda}{2} d}$$

where $\alpha = (2/\hbar)(2me)^{1/2}$, and where in the numerical formula, we have taken $\epsilon \approx 2$, the metallic value $\rho \ell \approx 10^{-11} \text{ ohm-cm}^2$, and d is in \AA and V_b in volts. For a given V_b , this has its minimum when $d = .9 V_b^{1/2}$, at which point

$$(RC)_{\min} \approx 6 \times 10^{-15} \text{ sec } V_b^{1/2}$$

But the tunneling approximation is only valid if V_b is large compared to kT , so that RC can typically not be reduced below 10^{-15} sec in a room temperature device. Still, this would be very fast and no serious problem till $\lambda \lesssim 1 \mu\text{m}$. However, the larger $\rho \ell$ value for a degenerate semiconductor might raise this to $\lambda \approx 10 \mu\text{m}$ for a Schottky. Thus, the RC limit leads to results quite similar to those described by (26).

IV. Quantum Limits

In addition to the classical RC limitations on high frequency performance of barrier diodes and the specific limit we found for the Schottky tunnel diode, there is a general

intrinsic quantum limitation. The essential point is that the quantum energy $\hbar\omega$ of the process provides a lower limit to the energy scale on which curvature can be effective. This in turn limits the effective S value of a tunnel diode to a quantum limit:

$$S_{\text{quantum limit}} = 2e/\hbar\omega \quad (27)$$

Let us see how this comes about.

As first shown by Tien and Gordon¹, the effect of applying an ac potential difference $V_1 \cos \omega t$ as well as a dc potential V_0 between two equipotential metals separated by a barrier is to replace the time dependence $\psi_k(\vec{r})e^{-iE_k t/\hbar}$ of the quantum wave function by

$$\psi_k(\vec{r})e^{-\frac{i}{\hbar}[E_k t + \int_0^t dt' (eV_0 + V_1 \cos \omega t')]} \quad (28)$$

or

$$\psi_k(\vec{r}) \sum_{n=-\infty}^{\infty} J_n \left(\frac{eV_1}{\hbar\omega} \right) e^{-(i/\hbar)(E_k + eV_0 + n\hbar\omega)t}$$

Thus, the electron can be considered to have a probability amplitude $J_n(eV_1/\hbar\omega)$ of being in a stationary state of energy $n\hbar\omega$ relative to its energy in the absence of the ac field, and this component of its wavefunction has relative phase $e^{-in\omega t}$. From this work of Tien and Gordon it is easy to see that the dc I-V curve in the presence of V_1 should be

$$I_{dc}^{QM}(V_0, V_1, \hbar\omega) = \sum_{n=-\infty}^{\infty} J_n^2 \left(\frac{eV_1}{\hbar\omega} \right) I_{dc}(V_0 + n\hbar\omega/e) \quad (28)$$

For low frequencies, this result corresponds to the classical result of averaging over the amplitude V_1 , but for high frequencies, a discrete quantum effect results. If we restrict our attention to terms in V_1 and V_1^2 , in the spirit of our basic approximation (2), (28) becomes

$$I_{dc}^{QM} = I_{dc}(V_0) + \frac{1}{2}V_1^2 \left. \frac{\Delta^2 I_{dc}}{\Delta V^2} \right|_{\omega} \quad (29)$$

where

$$\left. \frac{\Delta^2 I_{dc}}{\Delta V^2} \right|_{\omega} = \frac{I_{dc}(V_0 + \hbar\omega/e) - 2I_{dc} + I_{dc}(V_0 - \hbar\omega/e)}{(\hbar\omega/e)^2} \quad (30)$$

is the finite difference generalization of d^2I/dV^2 , to which it reduces at low frequencies. From this result, it is clear that $\hbar\omega/e$ forms a lower bound on the scale on which curvature is fully effective.

To complete the conversion of the S factor to a quantum form, we also must find how the high frequency conductance $dI/dV|_{\omega}$ is modified. This was done in an important new piece of work² reported by Tucker at this conference. They worked out the completely general case of an arbitrary collection of frequencies and amplitudes applied to a tunnel junction. If they specialize to the fundamental frequency current response to a single ac voltage, they find

$$I_{\omega}^{QM} = \sum_{n=-\infty}^{\infty} J_n \left(\frac{eV_1}{\hbar\omega} \right) \left[J_{n+1}(eV_1/\hbar\omega) + J_{n-1}(eV_1/\hbar\omega) \right] I_{dc}(V_0 + n\hbar\omega/e) \quad (31)$$

Retaining only terms linear in V_1 , this reduces to

$$\frac{I_{QM}}{V_1} = \left. \frac{\Delta I}{\Delta V} \right|_{\omega} = \frac{I_{dc}(V_0 + \hbar\omega/e) - I_{dc}(V_0 - \hbar\omega/e)}{2\hbar\omega/e} \quad (32)$$

Combining (30) and (32), we see that quantum effects require the replacement of $S = (d^2 I / dV^2) / (dI / dV)$ by

$$S_{QM} = \frac{\Delta^2 I / \Delta V^2}{\Delta I / \Delta V} = \left[\frac{2e}{\hbar\omega} \right] \frac{I'(+) - I'(-)}{I'(+) + I'(-)} \quad (33)$$

where

$$I'(\pm) \equiv \pm (e/\hbar\omega) [I_{dc}(V_0 \pm \hbar\omega/e) - I_{dc}(V_0)] \quad (33a)$$

is the slope of the chord connecting the point on the original I-V curve at $V=V_0$ with that at $V=V_0 \pm \hbar\omega/e$. The expression (33) clearly reduces to the classical one if $\hbar\omega \rightarrow 0$, and it reduces to a quantum limit $\pm(2e/\hbar\omega)$ if the curvature is so great over the scale of $(\hbar\omega/e)$ that $I'(+) \gg I'(-)$, or vice versa.

Application of Theory to Exponential Characteristic:

These general qualitative conclusions can be made quantitative by consideration of any specific model. They are particularly simple for the exponential tunnel diode with characteristic approximated (for dc) by $I = I_0(e^{S_0 V} - 1)$, where we now use the notation S_0 to indicate explicitly the dc value of S . For this case

$$\frac{\Delta^2 I}{\Delta V^2} = \left[\frac{2e}{\hbar\omega} \right]^2 \sinh^2(S_0 \hbar\omega/2e) I_0 e^{S_0 V} \quad (34a)$$

and

$$\frac{\Delta I}{\Delta V} = \left(\frac{e}{\hbar\omega} \right) \sinh(S_o \hbar\omega/e) I_o e^{S_o V} \quad (34b)$$

Note that the quantum conductance ($\Delta I/\Delta V$) rises exponentially when $S_o \hbar\omega/e \gg 1$, indicating a falling device impedance at high frequencies. Combining (34a) and (34b), we have

$$S_{QM} = \frac{2e}{\hbar\omega} \tanh \frac{S_o \hbar\omega}{2e} \quad (35)$$

independent of V_o . Clearly this reduces properly to S_o for $S_o \hbar\omega/2e \ll 1$, and to the quantum limit $2e/\hbar\omega$ for $S_o \hbar\omega/2e \gg 1$. Although derived for a special case, the qualitative conclusion should be general. That is, there is a quantum limit on the non-linearity-bandwidth product

$$S_o (\Delta\omega)_{\max} \approx 2e/\hbar \quad (36)$$

where S_o is the low-frequency value, and $(\Delta\omega)_{\max}$ is the frequency at which $(S_{QM}/S_o)^2 \approx \tanh^2 1 = 0.58 \approx \frac{1}{2}$. Note that we obtained exactly this same result earlier as (26), by an entirely different consideration of a Schottky tunnel diode limited by the plasma frequency. This supports the notion that this result is more general than our derivation shows.

The result (36) suggests the analogy to a linear amplifier of transconductance g feeding a parallel RC load. The voltage gain is

$$G = \frac{gR}{1+i\omega RC}$$

so that $G(0)_{\omega_{\max}} = (gR)(RC)^{-1} = g/C$, independent of R . That

is, one can raise the bandwidth of the flat response only by accepting lower gain at low frequencies. By adding inductance L in parallel, one can shift the peak response to $\omega_o = (LC)^{-\frac{1}{2}}$, but still one has

$$G_{\max} \Delta\omega = g/C \quad (37)$$

where $\Delta\omega$ is the half-power width.

If we apply this analogy to the non-linear case, we identify the maximum frequency $\omega_{\max} = 2e/\hbar S_o$ with $1/RC$. If we then note that for the diode $R = (dI/dV)^{-1} = S/(d^2I/dV^2)$, we can characterize the high frequency roll off in the non-linear response by an effective shunt capacitance

$$C_{\text{eff}} = \frac{\hbar}{2e} \frac{d^2I}{dV^2} \quad (38)$$

in parallel with the "electromagnetic" capacitance of the tunnel junction. Although this picture adds nothing new, it may provide a convenient way to model the quantum corrections at high frequencies. However, the analogy may not be quantitatively reliable, and the more general quantum scaling of all voltages, currents, and impedances described by Tucker and Millea may be required. They show that

$$I_{\omega}^* (V_o, V_1, V_2, \dots) = I_{\omega}^{\text{classical}} (V_o, V_1^*, V_2^*, \dots) \quad (39)$$

where

$$I_{\omega}^* = \frac{I_{\omega}}{\cosh(\hbar\omega/2kT)} \quad (40a)$$

$$v_i^* = v_i \frac{\sinh(\hbar\omega/2kT)}{(\hbar\omega/2kT)} \quad (40b)$$

and presumably this should hold for any exponential function in which kT is replaced by S/e .

In concluding this presentation of the Tucker theory of quantum corrections, it is appropriate to recall some of the limitations of the approach. Most important, it is derived for an ideal tunnel junction, in which the electric field exists only in the barrier. Thus it does not include true photo-emission in the electrodes, only the effect of the photons on the tunneling process itself. By the same token, it treats the tunneling process as an indivisible transfer of an electron from one side to the other. Hence, effects of scattering in the barrier, or junctions operating by thermal diffusion of electrons from one side to the other are not included in this present framework. None the less, it seems likely that the qualitative aspects of the non-linearity-bandwidth product limitation will apply quite generally.

Example: Super-Schottky at $T = 0$

As a test of this generality, consider the I-V of a super-Schottky at $T=0$, where there are no exponential tails. Then, from (24)

$$\begin{aligned} |I| &= \frac{1}{R} [v^2 - (\Delta/e)^2]^{\frac{1}{2}} & |v| &> \Delta/e \\ &= 0 & |v| &< \Delta/e \end{aligned} \quad (41)$$

where the sign of I is that of V . Inserting this in (33), with $eV_0 = \Delta$, we find

$$S_{QM} = \frac{2e}{\hbar\omega} \begin{cases} 1 & \Omega < 1 \\ \frac{(\Omega+1)^{\frac{1}{2}} - (\Omega-1)^{\frac{1}{2}}}{(\Omega+1)^{\frac{1}{2}} + (\Omega-1)^{\frac{1}{2}}} & \Omega > 1 \\ \frac{1}{2} \Omega^{-1} & \Omega \gg 1 \end{cases} \quad (42)$$

where $\Omega \equiv \hbar\omega/2\Delta$ is a normalized frequency. Thus the super-Schottky at $T=0$ has the full quantum-limit sensitivity $S_{QM} = 2e/\hbar\omega$ up to the gap energy $\hbar\omega = 2\Delta$, above which it falls below the quantum limit, eventually as $S_{QM} = 2e\Delta/(\hbar\omega)^2$. Note that this frequency dependence of S is rather different from (35), which points up the need to generalize (35) when the low frequency response is itself quantum limited, as it is here. One can get the low frequency behavior from (35) by setting $S_0 = \infty$, as is appropriate for a square corner, but this misses the further roll off above the gap frequency. This example illustrates the inadequacy of (35) when the I-V has more than one characteristic energy scale of importance, corresponding here to $S_0 = \infty$ (in the corner) and $S_0 \approx e/\Delta$ for the overall shape. In this case it appears that one gets the full quantum limit $S_{QM} = 2e/\hbar\omega$ (corresponding to $S_0 = \infty$) until the frequency corresponding to the lower value of S_0 ($\approx e/\Delta$) is reached. Presumably a more general formulation of these ideas could be established.

V. Concluding Discussion

For square-law or photon-counting detectors, the quantum limit (27) corresponds simply to the detectability of a single photon. But because of its wide-band sensitivity, such an IR detector is easily swamped by photons of background radiation distributed over the whole thermal spectrum. Thus the sensitivity limit in practice is commonly background limited rather than detector limited, and optimization depends critically on the specific application. For this reason, we shall now restrict our attention to heterodyne detectors.

Ideal heterodyne detectors are sensitive only to thermal photon noise in an IR bandwidth set by the bandwidth B of the i.f. amplifier following the mixer. For this to be the case, however, the local oscillator power must dominate the thermal noise power. For a mixer fed by a transmission line from an antenna, the thermal noise power is roughly kT per Hz over a band from $\nu = 0$ to $\nu = kT/h$, for a total of order $(kT)^2/h \approx 10^{-7} \text{W}$ at room temperature. Thus, only a modest P_{LO} is required to raise the sensitivity to the desired signal above the broadband background photon noise. Of course the system necessarily remains sensitive to thermal background radiation in the bandwidth of interest, namely

$$P_{th} = \frac{h\nu B}{e^{h\nu/kT} - 1} \quad (43)$$

which reduces to kTB for $h\nu \ll kT$, and to $h\nu B e^{-h\nu/kT}$ for

$h\nu \gg kT$. In a system with efficient radiation coupling the appropriate temperature to use in (43) will be determined primarily by the ambient seen by the antenna, rather than by the temperature of the detector element itself. However, if the mixer is resistive, it will itself contribute noise power corresponding to (43) but with ν referring to the i.f. frequency and T to the mixer temperature. Since kT at room temperature equals $h\nu$ for a photon of $\lambda \approx 50\mu\text{m}$, it is clear that IR detectors operate in a cross-over region. For example, for $\lambda = 10\mu\text{m}$, (43) shows that P_{th} is reduced by $\sim 5e^{-5} \sim 0.03$ relative to kTB . Thus, thermal noise in a room temperature mixer would greatly exceed the black-body noise in the input. But there is a well-known quantum limit on the minimum detectable power in a phase-sensitive (heterodyne) system

$$P_{\text{min,het}} = h\nu B \quad (44)$$

So long as $h\nu > kT$ for room temperature, then the thermal noise power due to incident room temperature radiation and to Johnson noise in the mixer are both below this quantum limit. Thus, in principle, it should be possible to approach the quantum limit of sensitivity for $\lambda \lesssim 25\mu\text{m}$ with a room-temperature heterodyne system, so long as the mixer conversion efficiency approaches $\frac{1}{2}$, and so long as the mixer introduces no more than Johnson noise.

We now turn to the problem of reaching the limiting conversion efficiency of $\frac{1}{2}$. From (16), the conversion effi-

ciency $\eta = 1/8 P_{LO} R_D S^2$. Thus, in principle, one can always reach $\eta = \frac{1}{2}$ by taking

$$P_{LO} = \frac{4}{R_D S^2}$$

In practice, for ease in impedance matching, R_D should be within an order of magnitude of the impedance of free space, $Z_0 = 377\Omega$. Thus a large value of S reduces P_{LO} , and with it, problems of heating and excess noise from the oscillator. Taking account of the quantum limit $S \leq 2e/\hbar\omega$, we find

$$P_{LO} \geq \frac{(\hbar\omega/e)^2}{R_D} = \frac{1.5 \text{ Watts}}{R_D (\Omega) \lambda^2 (\mu\text{m})}$$

For $\lambda = 10\mu\text{m}$, and $R_D = Z_0 = 377\Omega$, this is 4×10^{-5} Watts, which is still very small in terms of heating and other problems. (This P_{LO} could be reduced further for $R_D > Z_0$, but only by about two orders of magnitude before running into the minimum needed to override the thermal background.) Although heating should be no problem at $10\mu\text{m}$ with S equal to the quantum limit, it will inevitably be more of a problem with lower values of S and at shorter wavelengths. In fact, since dimensions must be kept below $\sim(\lambda/10)$, heating problems will generally scale up as λ^{-3} .

Since the quantum limit seems to raise no serious problems in P_{LO} , the problem is reduced back to finding an actual system which has an S approaching this quantum limit $S = 2e/\hbar\omega$ ($= 16\text{V}^{-1}$ at $10\mu\text{m}$). According to (35), a thermal

electron diode with $S_0 = 40 \text{ V}^{-1}$ would give $S = 15.8 \text{ V}^{-1}$, only one percent below the quantum limit. Also, for the computed tunnel Schottky $S_{dc} = 24 \text{ V}^{-1}$ estimated by (22) for $N_d = 5 \times 10^{18}$, (35) gives $S = 14.5 \text{ V}^{-1}$, only 10 percent below the quantum limit. On the other hand, the typical M-O-M has $S \approx 0.5 \text{ V}^{-1}$, which would require $(16/.5)^2 \approx 10^3$ times more P_{LO} to reach $\eta = \frac{1}{2}$. Since such large P_{LO} is beyond burnout, and introduces extra noise, ideal performance has not been approached with the MOM diodes.

Since it appears that S should be no problem in a semiconductor device for which $\omega_p \approx \omega$, this reduces the problem to optimizing the impedance level for matching the non-linear device into the radiation without excess loss in spreading resistance. The matching is simplified by working with R_D near the radiation resistance of an antenna $\sim 100\Omega$. To see whether this resistance level is reasonable, we recall that (25) implies for Schottky tunnel diodes that

$$R \approx \frac{\rho l}{A} e^{2eV_b/\hbar\omega_p}$$

where ρl is taken to be the geometric mean of the values for the metal and semiconductor, i.e., $\rho l \approx 2 \times 10^{-10} \text{ ohm-cm}^2$. To work at $10\mu\text{m}$, the area $A \lesssim (\lambda/10)^2 \approx 10^{-8} \text{ cm}^2$. For our design impedance of $\sim 100\Omega$ and $\omega_p \approx \omega$ for $10\mu\text{m}$, $V_b = 0.5$ volts, which seems not unreasonable, and a high doping level $\sim 5 \times 10^{18} \text{ cm}^{-3}$. With reasonable dimensions, it should be possible to hold the

spreading resistance down to $\sim R_D/10$, so that it does not degrade performance too much. But very careful attention to design of materials and geometry will be required if nearly limiting performance is to be obtained in this way.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

References

1. P. K. Tien and J. P. Gordon, Phys. Rev. 129, 647 (1963).
2. J. R. Tucker and M. F. Millea, preprint.

OPTICAL BAND GAPS OF OXIDES AND SULPHIDES

J. J. Gilman

ABSTRACT

Available data for the optical band gaps of oxides and sulphides are surveyed and shown to exhibit periodic behavior as a function of atomic number. No simple general correlations with other atomic properties could be found; including polarizabilities, electron affinities, ionization potentials, and interionic spacings. For the alkaline earth oxides taken alone, the gaps correlate well with the cation ionization potentials, and the inverse squares of the interionic distances.

OPTICAL BAND GAPS OF OXIDES AND SULPHIDES

J. J. Gilman

Currently, at least two emerging optical applications create a demand for oxide (or sulphide) materials. These are UV-transparent windows and the photoelectrolysis of water. The former needs materials with large band gaps (>3.5 eV.) and the latter requires anodically stable materials with relatively small band gaps (~1.6 eV.). Therefore, a survey of available data was made and the data were organized to exhibit systematic behavior. The principle sources of data for the oxides were the AIP Handbook¹, the Oxide Handbook², and the compilation by Strehlow and Cook³. They are arranged on a periodic chart in Figure 1. A similar periodic chart for the sulphides³ is shown in Figure 2. In both cases, the values are given for the highest oxide or sulphide.

Certain patterns may be seen in the periodic arrangements of both Figures 1 and 2. First, for the simple metal cations of Columns II, III, IV and V of the table, the band gap decreases with increasing atomic number. Second, there is a tendency for the gap to decrease in moving from left to right across the table. These trends are the same, but opposite to the trends of the cation electronegativities and polarizabilities.

Figure 1. SELECTED OPTICAL BAND GAPS OF OXIDES
(values in eV)

H	Li	Be	10.6	He	B	C	N	O	F	Ne
					7.0					
					Al	Si	P	S	C	Ar
					8.7	8.5				
K	Ca	Sc	Ti	V	Mn	Fe	Co	Ni	Zn	Ga
7.0	6.0	3.0	2.3	1.7	3.7	2.3	0.5	3.7	2.0	3.3
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag
5.8	5.6	5.0	3.5	3.7					1.5	1.2
Cs	Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au
5.1	1.1	5.6	4.6	2.8	2.3				0.2	
Fr	Ra	Ac								
			Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb
			2.7	0.8	1.0			1.2	1.7	1.4
			Th	Pa	U	Np	Pu	Am	Cm	Bk
			5.8					2.2		Cf
									Es	Fm
									Md	No
									Lw	

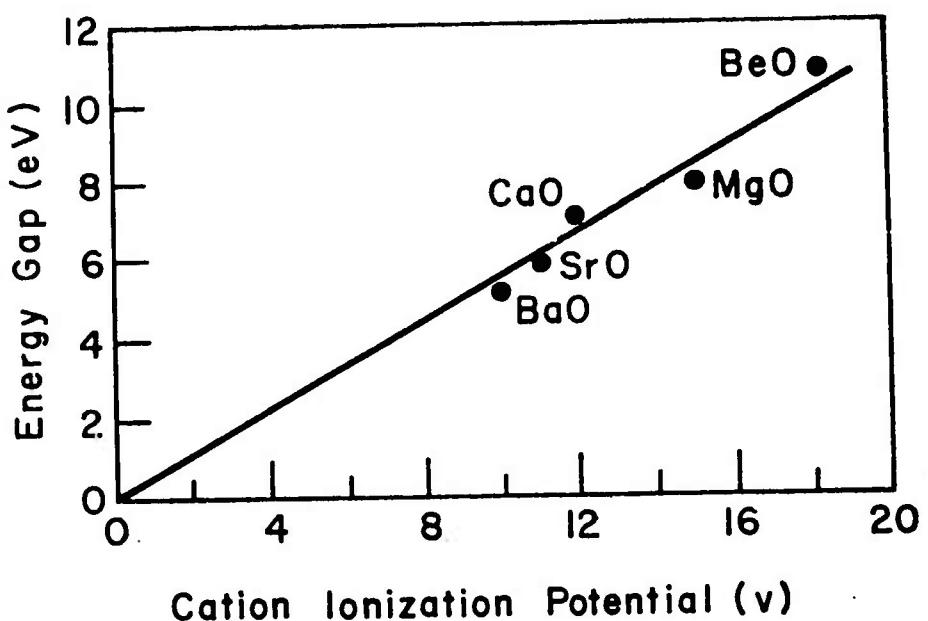
Figure 2. SELECTED OPTICAL BAND GAPS OF SULFIDES
(values in eV)

H		He		Li		Be		4.2		Na		Mg		2.4		K		Ca		Sc		Ti		V		Cr		Mn		Fe		Co		Ni		Cu		Zn		Ga		Ge		As		Se		Br		Kr	
Fr		Ra		Ac		Ce		Pr		Nd		Pm		Sm		Eu		Gd		Tb		Dy		Ho		Er		Tm		Yb		Lu																			
Th		Pa		U		Np		Pu		Am		Cm		Bk		Cf		Es		Fm		Md		No		Lw		Q33																							
Li		K		Ca		Sc		Ti		V		Cr		Mn		Fe		Co		Ni		Cu		Zn		Ga		Ge		As		Se		Br		Kr															
Na		Rb		Sr		Y		Zr		Nb		Mo		Tc		Ru		Rh		Pd		Ag		Cd		In		Sn		Sb		Te		I		Xe															
K		Rb		Sr		Cs		Ba		La		Hf		Ta		W		Re		Os		Ir		Pt		Au		Hg		Tl		Pb		Bi		Po		At		Rn											
2.1		2.1		4.8		3.9		1.3		2.9		0.1		1.1		1.1		2.0		0.9		0.8		2.0		1.4		0.41		1.2																					
2.0		1.2		0.9		6.2		1.2		0.12		1.2		3.9		3.6		3.5		2.4		1.9																													
5.8		1.7		0.12		1.9		1.8		1.5		0.75		1.0		2.4		2.3		2.1		1.7																													
4.8		1.7		0.12		1.9		1.8		1.5		0.75		1.0		2.4		2.3		2.1		1.7																													
2.4		1.3		2.9		0.1		1.1		1.1		2.0		0.9		0.8		2.0		1.4		0.41		1.2																											

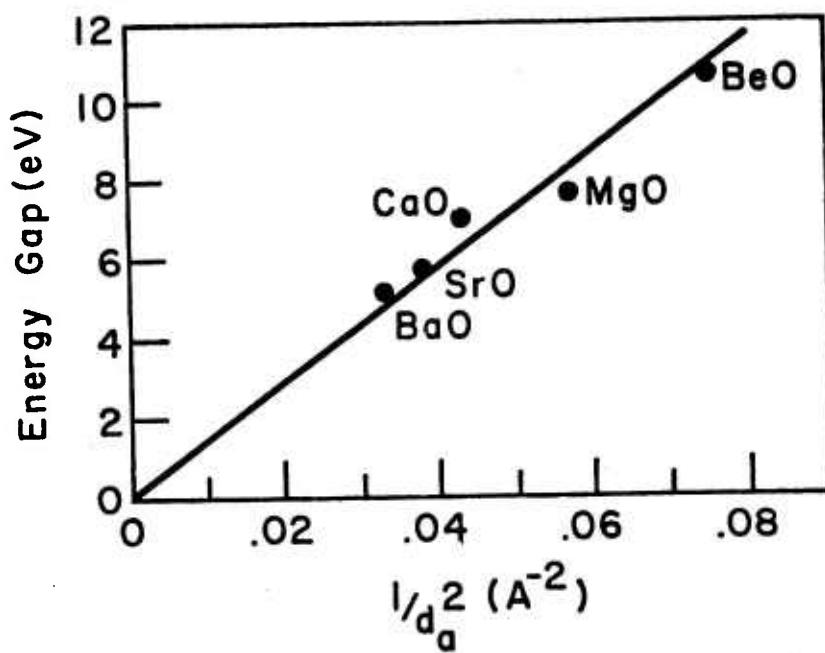
The gap value for B_2O_3 , does not fit the pattern and therefore may be anomalous. A value of 10-11 eV. would fit the pattern better and might make boron oxide a useful material for UV optical components. Crystals of B_2O_3 , can be readily grown using standard techniques, but apparently their optical properties have not been measured.

Figure 1 indicates that few binary oxides have energy gaps less than 2 eV. This does not include ternary oxides such as perovskites and tungsten bronzes. Sulphides, on the other hand, have smaller gaps in general as indicated by Figure 2; and many of them are smaller than 2 eV. Oxsulphides may give further opportunities for adjusting the band gap. Solid solutions formed between oxides with small and large gaps may do likewise.

Various simple correlations were investigated, but none with significant generality was found. These involve the radii, ionization potentials, electron affinities, and polarizabilities of the cations. For the most nearly fully ionic oxides (alkaline earth oxides) good correlations (Figure 3) exist between the energy gaps and the ionization potentials as well as the interionic spacings⁴. Such correlations for the other simple metal oxides are not good, probably because they have substantial covalent binding. For the transition metal oxides intra-ion excitations determine the gaps so simple correlations cannot be expected.



A



B

Figure 3. Correlations of Energy Gaps of Alkaline Earth Oxides:

- A - with ionization potential
- B - with inverse square of average interatomic distance.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC-71-C-0253 with The University of Michigan.

REFERENCES

1. D. E. Gray (ed.) "American Institute of Physics Handbook," 3rd Edition, McGraw-Hill, New York (1972) pp. 9-17.
2. G. V. Samsonov (ed.) "The Oxide Handbook," IFI/Plenum, New York (1973) p. 307.
3. W. H. Strehlow and E. L. Cook, "Compilation of Energy Band Gaps in Elemental and Binary Compound Semiconductors and Insulators," J. Phys. and Chem. Ref. Data 2, 163 (1973).
4. S. T. Pantelides, "Universal Valence Bands for Rocksalt-Type Compounds," Phys. Rev. B11, 5082 (1975).

WEAR WORKSHOP

A. Bienenstock, B. Budiansky, D. C. Drucker,
J. P. Hirth, H. Reiss, J. R. Rice

I. INTRODUCTION

A workshop on wear was held on July 15, 16 and 17, 1975. Presentations on various aspects of the subject were made by the following individuals invited to La Jolla for this purpose:

A. H. Shabaik (UCLA)
E. Rabinowicz (MIT)
N. P. Suh (MIT)
V. C. Westcott (Trans-sonics, Inc., Waltham)
F. Ling (RPI)
J. W. Butler (USNRL)
I. Finnie (Berkeley)

An overview of these presentations follows. Shabaik introduced the audience to the field by describing various categorizations of the wear process that have come to be conventionally recognized. Abrasive wear implies a ploughing action by hard, possibly loose, particles; fretting wear involves oscillatory motion between contacting surfaces; erosive wear refers to damage to and material removed from a surface by the impact of impinging solid or fluid particles, or a fluid stream; and corrosive wear suggests that the chemical effects of a hostile environment play a major damaging role in the wear process. In a somewhat different category is the widely used term adhesive wear, which embodies the implicit assumption that adhesion between parts of two rubbing surfaces occurs repeatedly

during wear. Rabinowicz outlined several of the somewhat loosely connected ideas that people in the field refer to, collectively, as the "adhesive theory of wear". In this "theory" wear is supposed to occur as the result of the adhesion and subsequent breaking off of opposing surface asperities, with the consequent formation of wear particles that are roughly equi-axed in shape. Suh presented his recent research on his "delamination theory of wear", which involves the removal of material in the form of thin flakes. Westcott showed the results of experimental studies of wear debris. Ling discussed the foundations of surface mechanics. Butler presented recent experimental information on the effects of surface ionization on wear rates. Finnie outlined his analysis of the process of erosive wear by solid particles.

Although the presentations were all very informative, primary attention by the MRC participants rapidly centered on the new ideas contained in Suh's delamination theory and their various conflicts with those of the older, so-called adhesive theory. (It should be noted that adhesion also plays a role in Suh's model; the terms delamination theory and adhesive theory will nevertheless be retained herein as a convenient, if imprecise, shorthand.) In part, this focus of interest occurred as a result of the explicit request by ARPA that the MRC study the delamination theory and suggest experiments to assess its degree and domain of validity. However, on its own merits as a provocative new set of ideas and point of depar-

ture for future research, Suh's delamination theory clearly deserved the serious attention it received.

There now follow brief outlines of the conventional adhesive theory, the delamination theory, and their points of conflict.

II. ADHESIVE THEORY OF WEAR

The conventional recital of the adhesive mechanism of wear goes something like this: A metal surface contains many asperities (bumps) which make contact with similar asperities on an opposing surface when the two surfaces are pressed together. When the surfaces are then slid past each other, a fraction of these asperities will be broken off (together with some surrounding material) as the result, typically, of the adhesion of two opposing asperities and the subsequent fracture of one of them. The fractured asperity may then remain a captive of the surface to which it adhered; it may break off as a free particle; it may be passed back and forth several times between the two surfaces before it finally breaks free. Let k be the probability that during an asperity-to-asperity encounter, one asperity becomes a free particle. Then, by a rough calculation due to Archard (1953), the volume V is wear material removed by sliding through a distance S is

$$V = \frac{kLS}{3H} \quad (1)$$

where L is the normal force and H is the hardness of the

softer material. Archard's calculation involved the assumption of a roughly equi-axed wear particle. The experts seem to accept the validity of (1) insofar as proportionality of V to L and S is concerned, although there does not appear to be a large reservoir of data to back this up. (Clearly, (1) does not include the influence of frictional heating, which becomes important in high wear regimes.) But no ab initio estimates of k are known, nor do there appear to exist any rational methods for its calculation. A superficial view of (1) would suggest that higher hardness reduces wear. But this is not always observed, and, of course, it can always be claimed by proponents of Archard's formula that k and H are not necessarily decoupled.

Different regimes of wear have been linked to experimentally determined values of k back-calculated from Eq. (1). Moderate to severe wear is considered to correspond to values of $k \approx 10^{-3}$ - 10^{-4} ; "benign" wear, during which machinery is considered operable over long periods of routine use, corresponds to $k \approx 10^{-7}$ - 10^{-8} .

III. DELAMINATION THEORY OF WEAR

Suh's delamination theory, like the adhesive theory, is really a collection of loosely connected ideas (see Suh's report in this volume) but (again like the adhesive theory) it does have a central concept, that is in striking contrast to that of the adhesive theory. Rejecting the model of wear

resulting from the knocking-off of asperities, Suh proposes instead that benign-to-moderate wear results from the formation of thin, platelike particles that flake off from the surface. These particles occur as a consequence of the nucleation, growth, and ultimate penetration to the surface, of subsurface cracks oriented nearly parallel to the surface. Suh presents striking microscopic pictures that display such platelike particles in various stages of formation, and is supported by the extensive observations by Westcott of wear debris containing particles in flake form.

Coupled with his observations, and the bold statement that flakes rather than chunks are primarily involved in the wear process, Suh also offers theoretical ideas to describe, at least qualitatively, the wear mechanism. He invokes the presence of a thin surface layer that is soft, nearly dislocation-free, and hence less liable to cracking than subsurface material. This surface layer is, he says, incapable of strain hardening (because dislocations escape to the surface), which is why it stays soft, even though its repeated deformation spreads it out so that it is everywhere in contact with the opposing surface.

IV. COMMENTS, SUGGESTIONS, AND RECOMMENDATIONS

Mostly as an outgrowth of Suh's provocative ideas and the controversies they engendered, we present a number of thoughts, speculations, and specific recommendations concerning

the wear problem.

a. Wear particles

It is not often that we are presented with a clash of opinions as striking, important, and yet susceptible to relatively straightforward resolution, as that exhibited by the flake-versus-chunk disagreement concerning wear particles. Suh says that in benign-to-moderate wear most of the mass loss occurs in the form of flakes; Rabinowicz, a leading proponent of the adhesive theory, has held that chunks contribute to most of the mass loss (even if platelike particles may be numerically predominant). It is evident that controlled experiments in various regimes of wear, incorporating careful measurement of wear particle size, shape, and mass, are essential, and are strongly recommended, in order to begin to resolve such a fundamental question.

b. Coordinated experimentation

We have gotten the distinct impression that experimentation in the process of wear has suffered from a lack of coordination, completeness, and documentation that has hindered progress in the achievement of understanding. Comparisons of results among different investigators is hindered by lack of uniformity of normal load, sliding speed and lubrication conditions, not to mention test apparatus and material specification and preparation. We have been distressed to find that such a basic quantity as the friction factor often goes unmeasured (or unreported). We would hope that a more systematic

program of basic experimental research will be encouraged. The observation of wear particles, as discussed in (a) would also, of course, be a highly desirable complement to such studies.

c. Surface geometry

An unsolved - indeed, mostly unstated - problem is that of the theoretical and experimental determination of the spectrum of surface geometry during steady-state wear. It would appear plausible to expect that a stationary roughness spectrum would be generated by steady wear conditions, and that a determination of this spectrum is crucial to the understanding of the wear process. In this connection, it is clearly difficult to accept the hypothesis that wearing surfaces can become "perfectly" smooth and everywhere in contact; for the very process of wear (be it via flake or chunk removal) introduces roughness on some scale, which in turn feeds back into the wearing behavior, in a tightly coupled fashion. We recommend that, from the spectral viewpoint, increased attention be paid to the question of steady-state surface roughness.

d. Wear particle distribution

In a similar spirit, the statistical distribution of size and shape of wear particles, in various regimes of wear, should be studied. (This suggestion goes beyond the gross evaluation discussed in (a) of which types of particles carry the most mass.) The spectrum of wearing surface geometry and wear particle distribution are obviously related, and so these

two kinds of investigation should be complementary parts of the same experiment - and, possibly, of a single theory.

e. Experimental surface techniques

Some of the newest techniques for examining surfaces may be useful and may not have been applied to the investigation of wear. The development of some of these techniques may have been heavily supported by ARPA. We have in mind techniques such as LEED, PES, XPS, Auger techniques, ion back scattering, molecular beams, Mossbauer effect (especially with steel surfaces), soft X-ray diffraction, NMR, as well as older techniques involving physical adsorption.

The participation of good fundamental physicists in surface studies is highly desirable. It is possible that substantial advances could be made. As a long range project an investigation of the possibility of using superconductive phenomena or the theory of tunnel diodes as aids to measurement might be considered.

f. Surface free energy

The role of surface free energy in wear is not entirely clear, although it is possibly very important. In recent years considerable advance in the theory of the thermodynamics of surfaces has occurred. Some of the new knowledge available in this respect could be brought to bear on the problem.

g. Stress, fatigue, and fracture analysis

It is highly desirable that theoretical studies be made of the near-surface stress distributions in the vicinity of

asperities induced by the normal and tangential loads involved in a wearing configuration. There are some qualitative reasons to expect that peak crack inducing stresses of the kind postulated by Suh, below the surface and outside the asperities, may indeed be predictable on a theoretical basis. This is clearly something to be settled by calculation. If such stresses are indeed found to occur, they might then be coupled with appropriate microscopic fracture, crack growth, or fatigue criteria, and possibly lead to rational theoretical understanding and predictive capability of wear rates and their dependence on more basic material properties.

h. The surface layer question

Much of the discussion evoked by Suh's theory concerned his presumption of a dislocation-free, highly deformed, fracture resistant layer at the wearing metal surface. There is, first of all, confusion engendered by references of workers in the field to at least three kinds of surface layers; the amorphous Bilby layer, a "shear-mix" layer, and now Suh's soft dislocation-free layer. Verification of the existance of one or more of these layers and their physical descriptions is clearly desirable. Suh's "image forces" that remove dislocations can hardly be expected to be effective beyond a depth of a few hundred Angstroms. (This is not inconsistent with Suh's own views.) We note, however, that the soft, fracture-proof layer may not be at all a necessary part of the mechanism of flaking wear. If the speculations concerning the occurrence of peak stresses

below the surface are correct the integrity of the thin layer above the peak stress location is then explicable on the basis of low stress rather than high fracture strength. The same argument can also be used to explain Suh's success in reducing wear rates by means of a very thin overlay of soft metal. An inherently lower frictional coefficient of the overlay can suffice to lower the applied tangential stresses, raising the fatigue life of the substrate; and a sufficiently thin overlay would not be subjected to the higher subsurface stresses (assuming the validity of our speculation concerning their distribution). But all of this is speculative, and only points up further the need for sensible analysis and experiment. There may be a soft layer, and, if so, genesis should be determined; but the delamination theory, at present, need not be tied to the existence of this layer.

i. Wear diagnosis

From the immediate practical point of view, there are some methods available for identifying the source (by machine part) of wear particles (e.g., some of Westcott's techniques). This suggests the possibility that a part, about to fail, can be identified without dismantling the entire system. The chances for accomplishing this for some systems are judged to be good. Therefore additional support of work aimed at this goal is recommended. Again some of the newest techniques of fundamental physics and chemistry may be valuable in this respect.

j. Ion implantation

The results of experiments performed at NRL on the effects of ion implantation on wear rates are interesting, but they must be regarded as preliminary, if provocative. Coordinated experimentation, in the spirit of item (b) above, including measurements of friction, wear rate, etc., is essential in any further ion implantation tests. With this proviso, such tests, we believe, should be continued to the point where a sober, rational assessment of the effectiveness of ion implantation in reducing wear can be made.

v. CONCLUDING REMARKS

Wear between sliding parts of a machine is a serious national problem. Often the loss of a few grams, perhaps even a few milligrams of metal in a crucial part of the machine may be responsible for the abandonment of hundreds of pounds or even tons of machinery, or, at the very least, occasionally very expensive remachining procedures. The economic and conservational consequences of this are self-evident. In the face of this the community of workers engaged in the study of wear on a scientific basis, and the number of fundamental scientists involved, is tiny. Furthermore very little public funding is available for research and development enterprises in the field of wear.

The so-called delamination theory is one of the few new ideas in this field in some time. Irrespective of its ultimate

validity or, at least, of its universality, it has stimulated much new thinking and presented enough compelling visual and experimental evidence to demand serious attention as well as investigation. In this sense it represents a point of departure in the development of the theory of wear, and on this basis alone ARPA is advised to support more research in wear as a field which may be charged with potential at the moment, and one which could yield handsome rewards in the near future. Efforts should be made to increase the number of highly qualified scientists interested in this field.

ACKNOWLEDGEMENT

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

QUASI-STATIC SOLID PARTICLE DAMAGE
IN BRITTLE SOLIDS

A. G. Evans

ABSTRACT

Observations of radial and lateral fracture from plastic indentations have been used to examine several important consequences of indentation fracture; notably, fracture toughness determination by indentation, material removal by abrasive processes, and material removal by low velocity solid particle impact. Good correlations of the predicted behavior with available empirical data confirm the promise of this approach.

QUASI-STATIC SOLID PARTICLE DAMAGE
IN BRITTLE SOLIDS

A. G. Evans

I. INTRODUCTION

Quasi-static indentation studies in brittle materials have established that the indentation is plastic when the indenter radius decreases below a critical value, b_c , which depends on the hardness, $H^{1/2}$; and that the mode of fracture accompanying plastic indentation is very different from the fracture observed during elastic indentation.³ The fracture processes involved in elastic indentation have been extensively investigated, and related to phenomena such as solid particle erosion.¹ However, many of the practical phenomena that involve indentation are usually in the plastic regime, e.g., erosion due to small angular (dust) particles, abrasive wear, machining, hardness measurements. An analysis of the fracture created by plastic indentation is thus an important prerequisite for understanding such phenomena.

A recent classification of quasi-static indentation fracture,² in the radial (surface) and lateral (sub-surface) modes, provides a basis for evaluating the parameters that exert the prime influence on the fracture extension. The implications of these results for three important problems are

explored in this paper, notably: fracture toughness determinations from radial fracture measurements; material removal during abrasive wear (or sawing, or machining) by lateral crack extension; material removal during low velocity solid particle erosion, again by lateral crack extension.

The fracture toughness formulation can be derived directly from the radial crack extension relation, but the material removal estimates require the following additional information; an evaluation of the effect of adjacent indents on lateral crack extension and, for the erosion problem, a determination of the relation between the dynamic force and the projectile velocity. These problems are addressed before proceeding with the proposed discussion of indentation fracture implications.

II. INTERACTION BETWEEN ADJACENT INDENTS

An analytical evaluation of the interaction between adjacent indents can be performed if the stress field that controls the crack propagation behavior can be expressed in a relatively simple functional form. A form which adequately describes the observed lateral crack propagation (see Appendix I) is based on the analogous stress field for a pressurized spherical cavity⁴ (Eq. A12);

$$\sigma_{\theta} = \frac{\sigma_y}{3} \left(\frac{a}{r} \right)^3 \xi^3$$

where σ_{θ} is the tangential stress, σ_y is the uniaxial yield

stress, r is the distance from the center of the impression, a is the impression radius, and ξ is a constant (determined by plastic boundary location). The corresponding residual stress for a pressurized cavity, σ_θ^R , at the zero pressure condition is;⁴

$$\sigma_\theta^R = \frac{\sigma_Y}{3} \left(\frac{a}{r} \right)^3 [\xi^3 - 1 - 3 \ln \xi] \quad (1)$$

For two adjacent indents the total tangential stress, σ_θ^T , at any position between the pre-existing indent and the current indent, separated by r_o , is simply the sum of Eqs. (1) and (A12);

$$\sigma_\theta^T = \frac{\sigma_Y}{3} \left(\frac{a\xi}{r} \right)^3 \left[1 + \left\{ 1 - \frac{1}{\xi^3} (1 + 3 \ln \xi) \right\} \left(\frac{r}{r_o - r} \right)^3 \right] \quad (2)$$

where r is now the distance from the center of the current impression. By considering an array of indents with circular symmetry, an approximate value for the stress intensity factor, K , at the current indentation can be obtained directly from Eq. (2) using the K formulation discussed in Appendix I;

$$\frac{K}{\sigma_Y} \approx \frac{2a^3 \xi^3}{3\sqrt{\pi}C} \int_a^C \frac{\{1 + \beta[r/(r_o - r)]^3\}}{r^2 \sqrt{C^2 - r^2}} dr \quad (3)$$

where $\beta = 1 - \xi^{-3} - 3\xi^{-3} \ln \xi$ and C is the lateral crack radius measured from the center of the impression. Integration of Eq. (3) gives;

$$\frac{K}{\sigma_Y \sqrt{a}} \cdot \left[\frac{3\sqrt{\pi}}{2\xi^3} \right] = \left(\frac{a}{C} \right)^{5/2} \left[\sqrt{(C/a)^2 - 1} + \left(\frac{C}{a} \right)^2 \left(\frac{a}{r_o} \right)^2 f \right] \quad (4)$$

where

$$f = \frac{\beta}{2} \left\{ \frac{[4 - (C/r_o)^2]}{[1 - (C/r_o)^2]^{5/2}} \left[\frac{\pi}{2} + \sin^{-1} \left[\frac{(C/r_o)^2 - (a/r_o)}{(C/r_o)[1 - (a/r_o)]} \right] \right] \right. \\ \left. + \frac{\sqrt{(C/a)^2 - 1}}{[(r_o/a) - 1]^2 [1 - (C/r_o)^2]^2} \frac{6 - 2(C/r_o)^2 (a/r_o) - 3(C/r_o)^2 - 5(C/r_o)}{6 - 2(C/r_o)^2 (a/r_o) - 3(C/r_o)^2 - 5(C/r_o)} \right\}$$

The relation between the left hand side of Eq. (4) and C/a is plotted in Fig. 1 for various r_o/C , taking $\xi = 1.1$ (see Appendix I). It is apparent that the residual stress around the pre-existing indent can significantly increase the stress intensity factor at the current impression. This is qualitatively consistent with the observations of lateral fracture initiation in zinc selenide,² and of the larger extension of lateral fracture in several harder materials,² under multiple indent conditions.

The critical indent separation for lateral crack linkage, r_o^c , is determined by the crack instability position, i.e., where K increases with increase in crack length. Such instabilities do not occur, however, at crack separations, where crack interaction effects are unimportant,⁵ i.e., less than the impression diameter. The detailed determination of r_o^c thus requires an analysis of crack interaction effects on K . This problem is complex, but the interaction term for two through-the-thickness cracks in a uniform stress field⁶ provides an estimate of the interaction magnitude;

$$R_K = [(r_o/C) - 1]^{\frac{1}{2}} \frac{[(r_o/C)^2 + 1] \left[\frac{Q_1(X)}{Q_2(X)} - 1 \right] + 2(r_o/C) \left[\frac{Q_1(X)}{Q_2(X)} + 1 \right]}{2(r_o/C)^{\frac{1}{2}} [(r_o/C) + 1]} \quad (5)$$

where $Q_1(X)$ and $Q_2(X)$ are the elliptic integrals;

$$Q_1(X) = \int_0^{\pi/2} \frac{d\phi}{\sqrt{1-X^2 \sin^2 \phi}}$$

$$Q_2(X) = \int_0^{\pi/2} \sqrt{1-X^2 \sin^2 \phi} d\phi$$

and R_K is the stress intensity factor ratio. The interaction modified stress intensity factors are plotted as the dotted lines in Fig. 1. Instability points are now apparent, for $r_o/a > 4$. The instabilities define r_o^c values, which are plotted in Fig. 2. For $r_o/a < 4$, K appears to increase continuously with increase in crack length and r_o^c in this range is probably initiation controlled, as indicated on Fig. 2. For $r_o/a > 4$, r_o^c is propagation controlled and the critical indent separation is uniquely characterized by the normalized variables, r_o^c/a and $K_c/\xi^3 \sigma_Y \sqrt{a}$.

The predicted indent separation for lateral crack linking can be compared with experimental data obtained for zinc sulphide and alumina² by using the coordinates r_o^c/a and $K_c/\sigma_Y \sqrt{a}$ (ξ is not readily determined by experiment in these materials²). A logarithmic plot of the data is shown in Fig. 3; also shown is the predicted curve for $\xi = 1.1$ (see Appendix

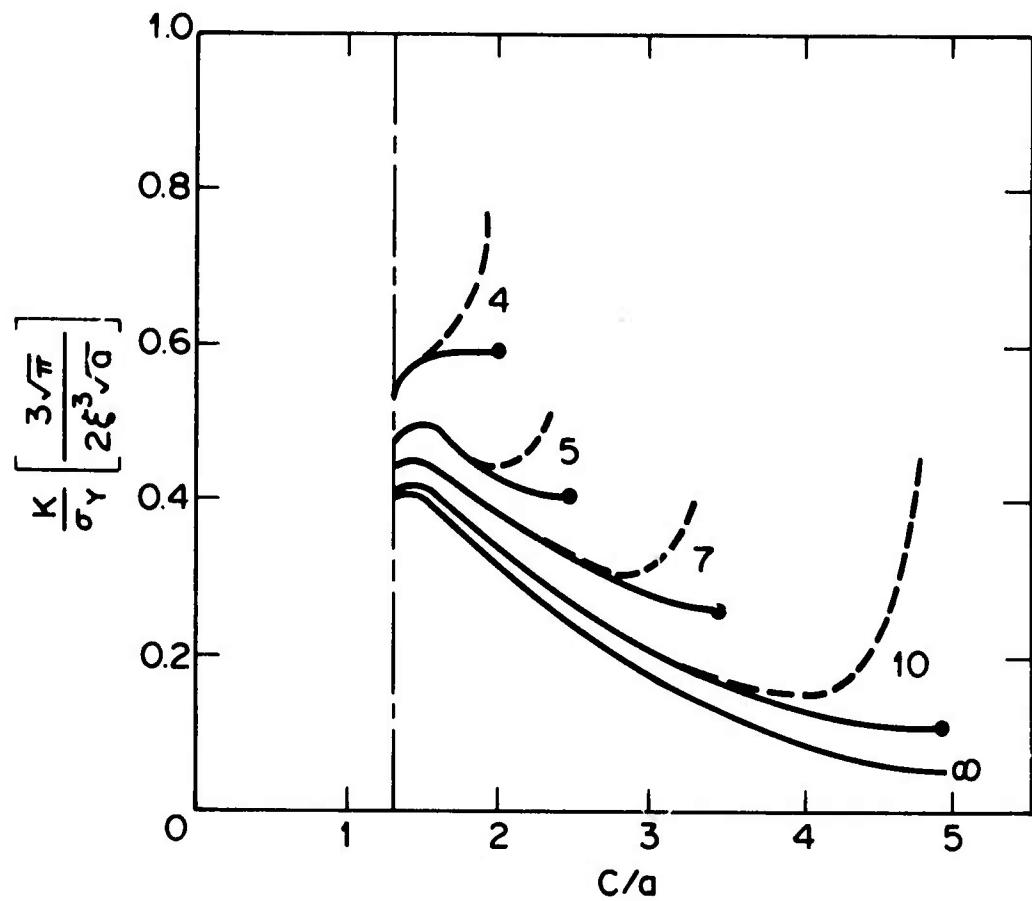


Figure 1. A plot of the normalized stress intensity factor $(K/\sigma_0)(3\sqrt{\pi}/2\xi^3\sqrt{a})$ versus the normalized lateral crack size C/a for an array of indents with circular symmetry; the numbers adjacent to each curve represent the values of the normalized indent separation r_0/a . The dotted lines are the approximate crack interaction modified stress intensity factors.

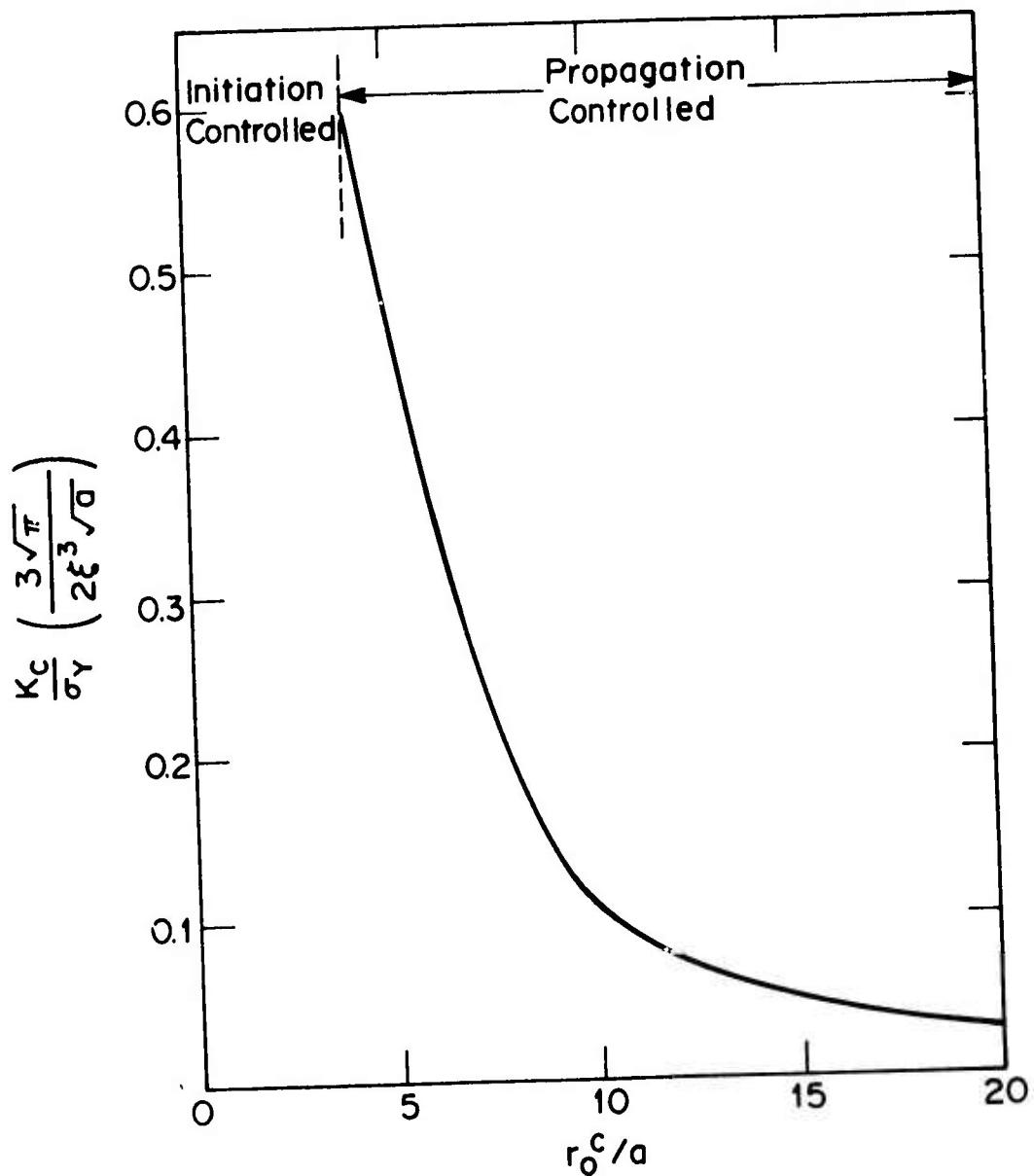


Figure 2. The variation of the normalized stress intensity factor with the normalized critical indent separation for lateral crack linkage.

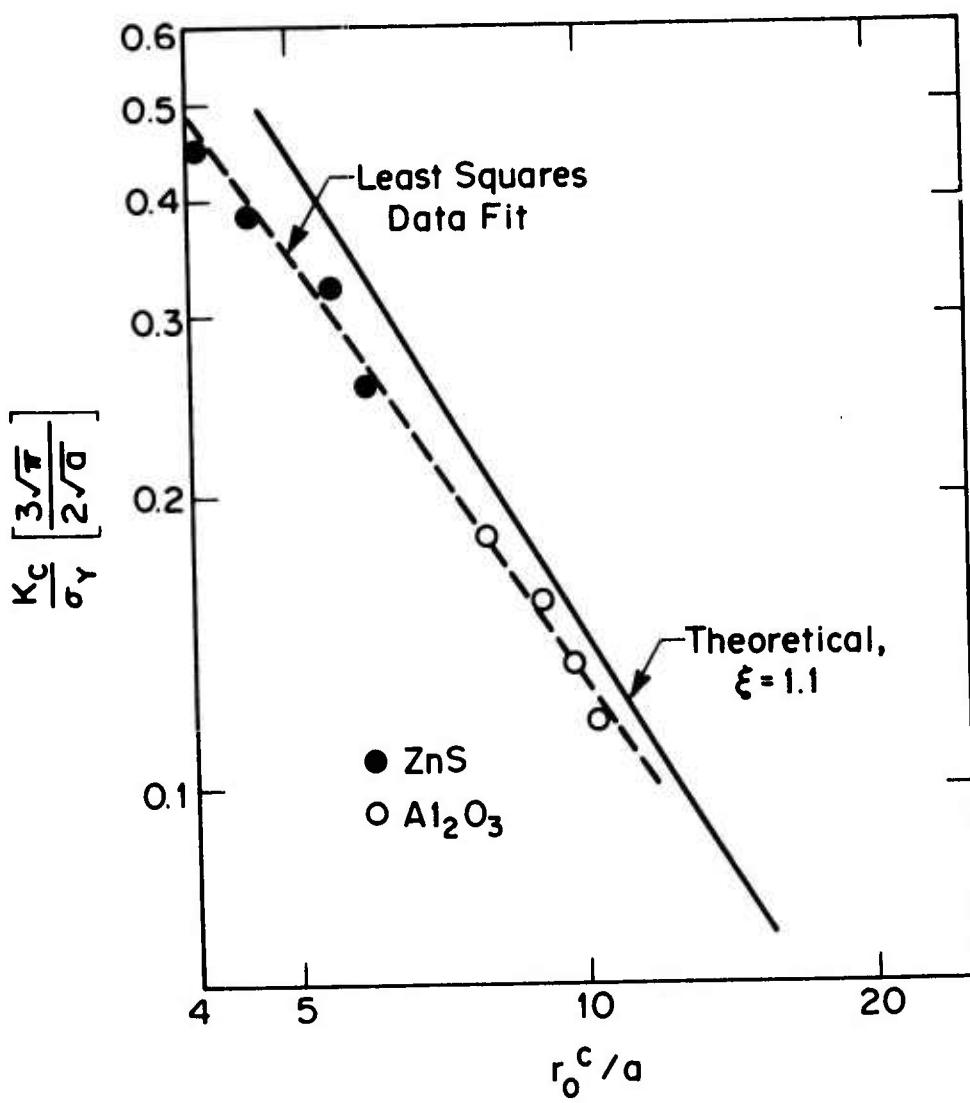


Figure 3. A logarithmic plot of the normalized stress intensity factor versus the normalized critical indent separation for zinc sulphide and alumina; also shown, by the full line, is the predicted result from Fig. 2 for $\xi = 1.1$.

I) taken from Fig. 2. The good correlation of the predicted behavior with experimental observation verifies the utility of the pressurized cavity analogy for assessing lateral fracture processes. This analogy may thus be used to examine other important lateral fracture phenomenon such as unloading effects (due to residual stress development)², dynamic extension under high velocity impact, etc.

A least squares fit of the test data (Fig. 3) yields the following functional relation for the critical indent separation;

$$r_o^c = 1.3 \left(\frac{\sigma_y}{K_c} \right)^{2/3} a^{4/3} \quad (6)$$

Comparison with the lateral crack extension for a single indent (Appendix II) gives;

$$\frac{r_o^c}{c} = 2.5 \left(\frac{K_c}{\sigma_y \sqrt{a}} \right)^{1/12} \quad (7)$$

This result, which applies provided that $K_c \gtrsim 10^{-4} \sigma_y \sqrt{a}$ (i.e., $r_o^c \gtrsim c$), shows that the interaction effect diminishes (albeit rather slowly) as the impression diameter (or the yield strength) increase, but increases as the fracture toughness increases.

III. THE IMPACT FORCE

The force, P , exerted by a spherical projectile when it impacts a target has been calculated by Timoshenko and Goodier⁷ for fully elastic contact conditions, and the result has been shown to apply with good validity up to appreciable projectile

velocities⁸ (e.g., contact times at least as small as 10^{-5} s). The forces developed when the contact is plastic can be calculated using a similar approach.

From the equation of motion;

$$mvdv = -Pdz \quad (8)$$

where m and v are respectively the mass and velocity of the projectile and z is the vertical displacement of the boundary, we see that a relation between P and z is needed for dynamic force calculations. There are no complete analytic solutions for $P(z)$, but two options are available for determining approximate functional expressions; the plastic displacement for a rigid indenter (determined directly by the indenter profile) can be combined with the approximate elastic displacements², or the numerical solution of Hardy et al⁹ can be described by one or more fitting functions. The former approach turns out to be unsatisfactory because it does not effectively account for the critical elastic displacement range, prior to plastic indentation. The latter approach provides reasonable results, and is thus described in some detail; although undoubtedly this result can be improved upon when more extensive numerical displacement solutions become available.

The rigid indenter solution of Hardy et al⁹ for a non-work hardening material, reproduced in Fig. 4, is adequately described in the range $P \leq 10P^*$ by the elastic (Hertzian) result⁷;

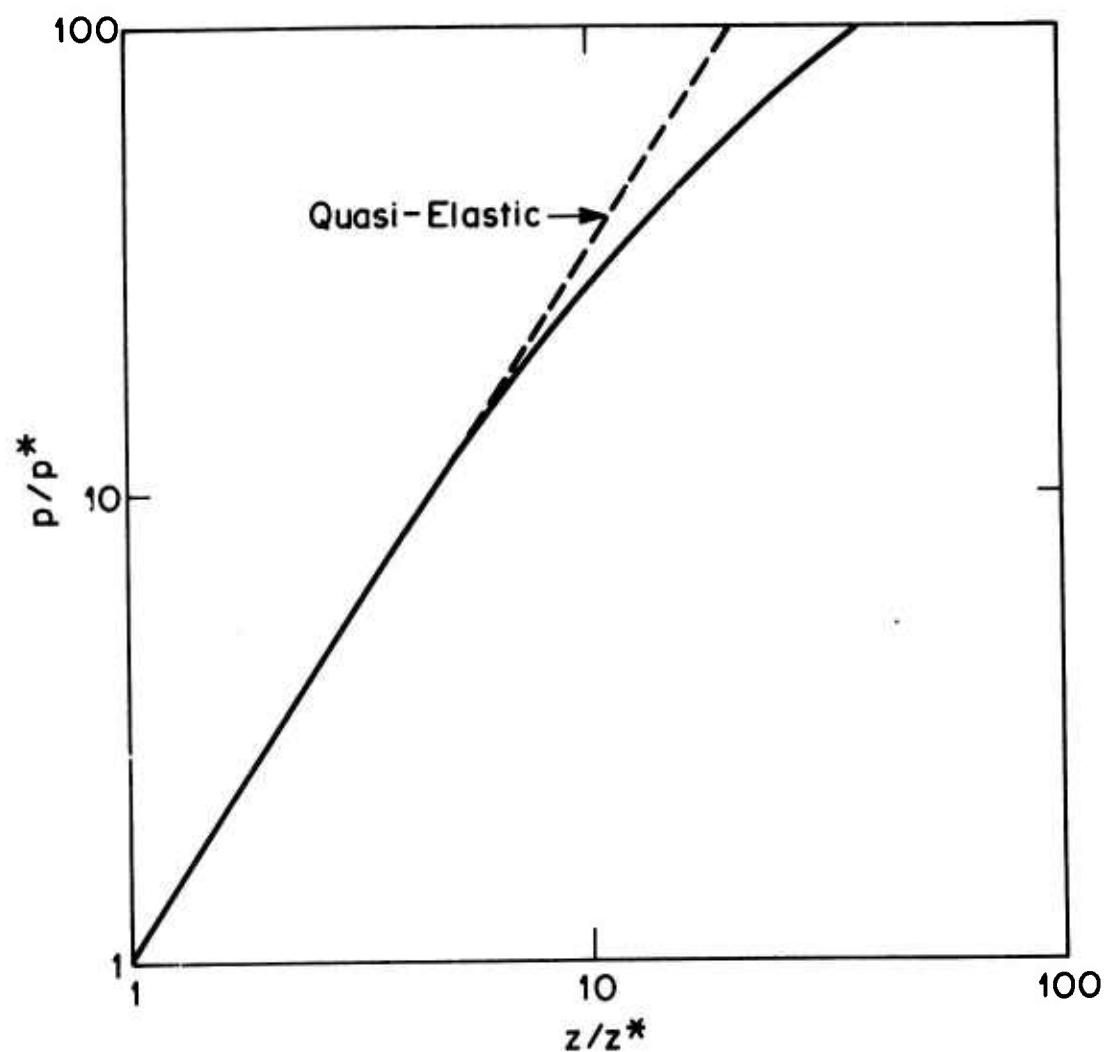


Figure 4. Force, displacement behavior for a rigid spherical indenter and a non-work hardening material (taken from Ref. 9).

$$P = 3.6 G^t \sqrt{b} z^{3/2} \quad (9)$$

where G^t is the shear modulus of the target, b is the projectile radius and P^* is given by⁹;

$$P^* = 1.8 \times 10^{-2} \frac{\sigma_Y^3 b^2}{(G^t)^2} \quad (10)$$

For $P > 10P^*$ the following expression provides a good fit to the numerical displacement curve;

$$\frac{P}{eP^*} = \ln^3 \left(\frac{z}{z^*} \right) \quad (11)$$

where z^* is given by;

$$z^* = 2.9 \times 10^{-2} b \left(\frac{\sigma_Y}{G^t} \right)^2 \quad (12)$$

Substituting the values for P from Eq. (9) and (11) into Eq. (8) and integrating to obtain the peak displacement, z^m , gives, for $z^m > 4.6 z^*$ ($P > 10P^*$);

$$\begin{aligned} 42 \left(\frac{v_i}{c_s^p} \right)^2 \left(\frac{G^p}{G^t} \right) \left(\frac{G^t}{\sigma_Y} \right)^3 - 0.45 \left(\frac{\sigma_Y}{G^t} \right)^2 \\ = \frac{z^m}{b} \left[\ln^3 \left(\psi \frac{z^m}{b} \right) - 3 \ln^2 \left(\psi \frac{z^m}{b} \right) + 6 \ln \left(\psi \frac{z^m}{b} \right) - 6 \right] \quad (13) \end{aligned}$$

$(z^m > 4.6 z^*)$

where $\psi = 34.5 (G^t/\sigma_Y)^2$, G^p is the shear modulus of the projectile[†] and c_s^p is the shear wave velocity in the projectile.

[†]The numerical displacement solution⁹ applies for a rigid indenter, Eq. (13) is thus only strictly valid when $G^p > G^t$.

For $z^m < 4.6 z^*$ ($P < 10P^*$) the elastic result applies, viz;⁷

$$\frac{z^m}{b} = 1.17 \left(\frac{v_i}{c_s^p} \right)^2 \left(\frac{G^p}{G^t} \right)^{2/5} \quad (14)$$

$(z^m < 4.6 z^*)$

Values of z^m/b can thus be computed as a function of $(v_i/c_s^p)^2$ (G^p/G^t) for various (σ_y/G^t) as shown in Fig. 5. The yield stress independent regime for small z^m/b is the quasi-elastic region ($z^m < 4.6 z^*$) determined by Eq. (14), and the yield stress dependent regime at larger z^m/b is the elastic/plastic region ($z^m > 4.6 z^*$) dominated by Eq. (13).

The dynamic force can be obtained using a similar procedure. For $P > 10P^*$;

$$e^{2.7(G^t/\sigma_y)(P^m/b^2G^t)^{1/3}} \left[21 \left(\frac{P^m}{b^2G^t} \right) \left(\frac{G^t}{\sigma_y} \right)^3 - 23 \left(\frac{P^m}{b^2G^t} \right)^{2/3} \left(\frac{G^t}{\sigma_y} \right)^2 \right. \\ \left. + 16 \left(\frac{P^m}{b^2G^t} \right)^{1/3} \left(\frac{G^t}{\sigma_y} \right) - 123 \right] \quad (15)$$

$$= 3 \times 10^3 \left(\frac{v_i}{c_s^p} \right)^2 \left(\frac{G^p}{G^t} \right) \left(\frac{G^t}{\sigma_y} \right)^5 - 7$$

For $P < 10P^*$;

$$\frac{P^m}{b^2G^t} = 4.6 \left[\left(\frac{v_i}{c_s^p} \right)^2 \left(\frac{G^p}{G^t} \right) \right]^{3/5} \quad (16)$$

The force can thus be plotted using the normalized coordinates (P^m/b^2G^t) and $(v_i/c_s^p)^2 (G^p/G^t)$ for various (σ_y/G^t) as shown in Fig. 6. It is apparent that the quasi-elastic result applies at low velocities, but at higher velocities deviations occur

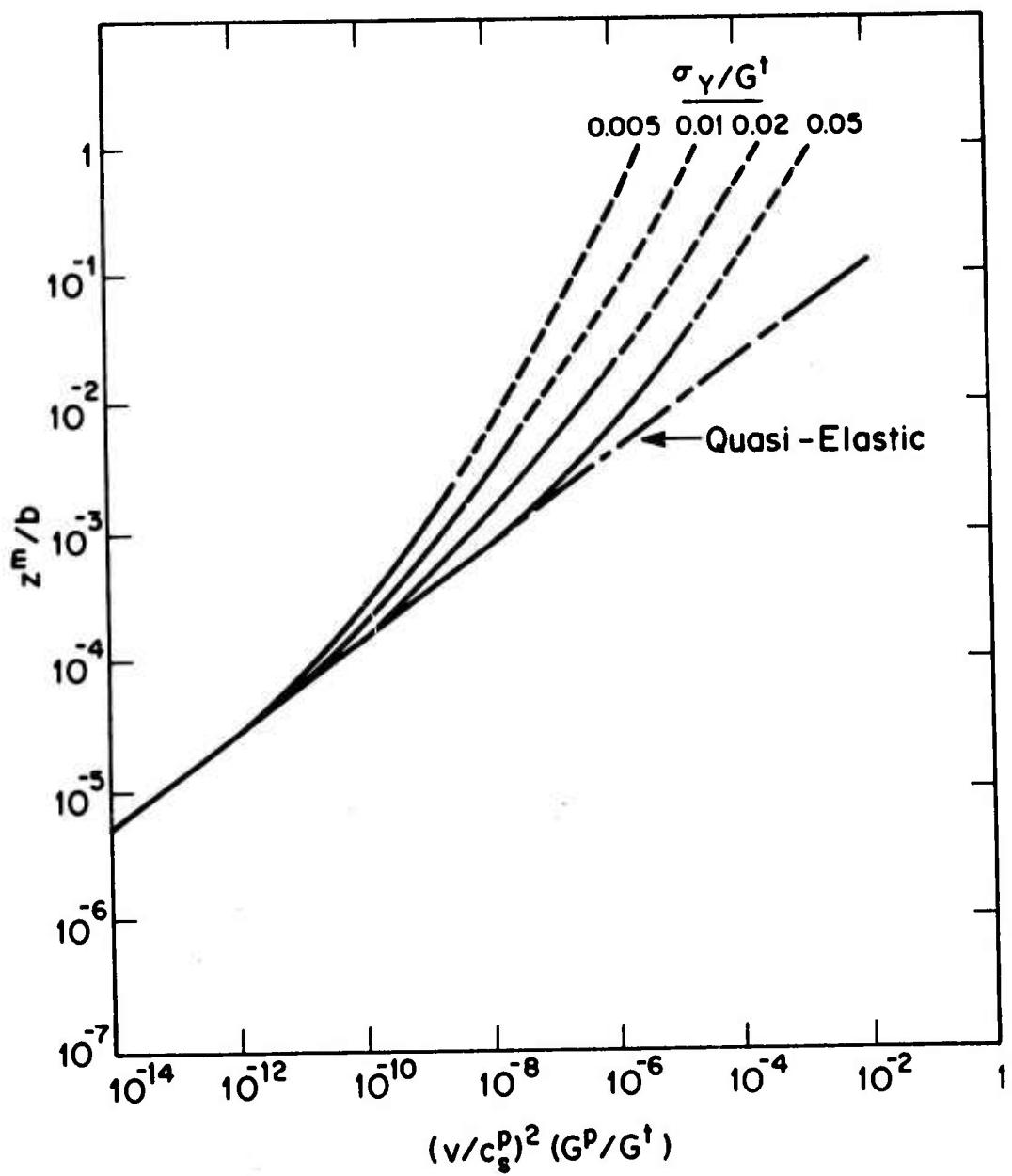


Figure 5. The peak displacement, z^m , for a sphere impacting a target with a velocity, v . The numbers adjacent to each curve are the corresponding values of the ratio σ_y/G^t .

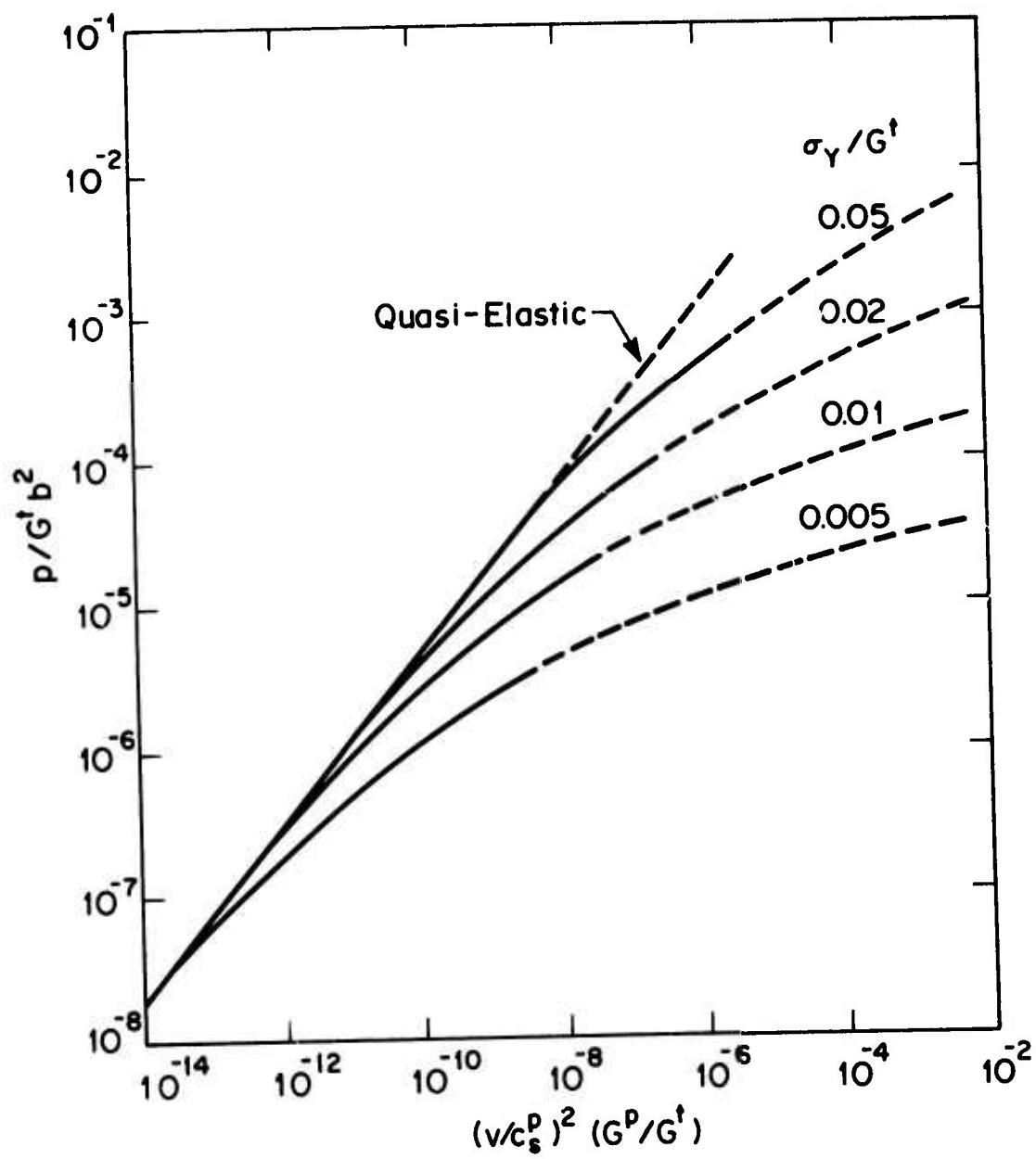


Figure 6. The peak force, velocity relation for various σ_y/G^t .

such that the force is less than that expected for fully elastic contact, with the deviations being larger for the lower yield strength materials.

The limitation of this force diagram is its exclusion of work-hardening. Significant work-hardening would tend to extend the effective range of the quasi-elastic regime and would generate forces larger than those incurred by a non-work hardening material. An estimate of the importance of work hardening can be made for fully plastic indentation, which should be a good approximation for very small σ_y/G^t . For a work hardening exponent, x , the force is related to the impression radius by²

$$P = \pi \delta a^{2+x} \quad (17)$$

where δ is a constant. For a fully plastic material the vertical displacement is given directly by the sphere profile.

$$z = b - \sqrt{b^2 - a^2} \quad (18)$$

Substituting for P and z from Eq. (17) and (18) into Eq. (8) and integrating gives;

$$2 \frac{(3+x)}{3} \left(\frac{G^p}{G^t} \right) \left(\frac{v}{c_s^p} \right)^2 = \frac{\delta a^x}{G^t} \left\{ \left(\frac{b}{a} \right)^x [\sin^{-1} (a/b)]^{2+x} - \left(\frac{a}{b} \right)^2 \sqrt{1 - (a/b)^2} \right. \\ \left. - \left(\frac{a}{b} \right)^x \frac{(x^2+3x+2)}{6(4+x)} [\sin^{-1} (a/b)]^{4+x} + \dots \right\} \quad (19)$$

Substituting for a from Eq. (17) gives the desired relation between P^m and v . For illustration purposes we shall consider the case, $a \gtrsim 0.3b$, whereupon Eq. (19) reduces to;

$$\left(\frac{G^P}{G^t}\right) \left(\frac{v}{c_p^p}\right)^2 = \frac{3(4-x^2)}{4(3+x)(4+x)\pi^2} \left(\frac{P^m}{G^t b^2}\right) \left(\frac{G^t}{H^*}\right)^2 \quad (20)$$

where H^* ($= \delta a^x$) is the apparent hardness ($\equiv P^m/\pi a^2$), which depends on the absolute size of the indenter, etc.² Hence, there is no single force diagram for materials which work harden. However, the force ratio, R_F (i.e., for work-hardening materials compared to non-work hardening materials with the same basic hardness), which can be derived directly from Eq.

(20), as;

$$R_F = \left[\frac{(3+x)(4+x)}{3(4-x^2)} \left(\frac{H^*}{H} \right) \right]^{\frac{1}{2}} \quad (21)$$

gives a measure of the importance of work hardening. For example, in zinc sulphide, x is 0.3 and H^*/H ranges between 1 and 2 for most practical situations²; thus $1.1 < R < 1.6$, indicating that the effects of work hardening are small but significant.

IV. IMPLICATIONS

A. Fracture Toughness Determinations

A characterization of radial fracture² shows that the radial crack length, C_r , can be described (Fig. 7) over a large range (0.6a to $\gtrsim 5a$) by:

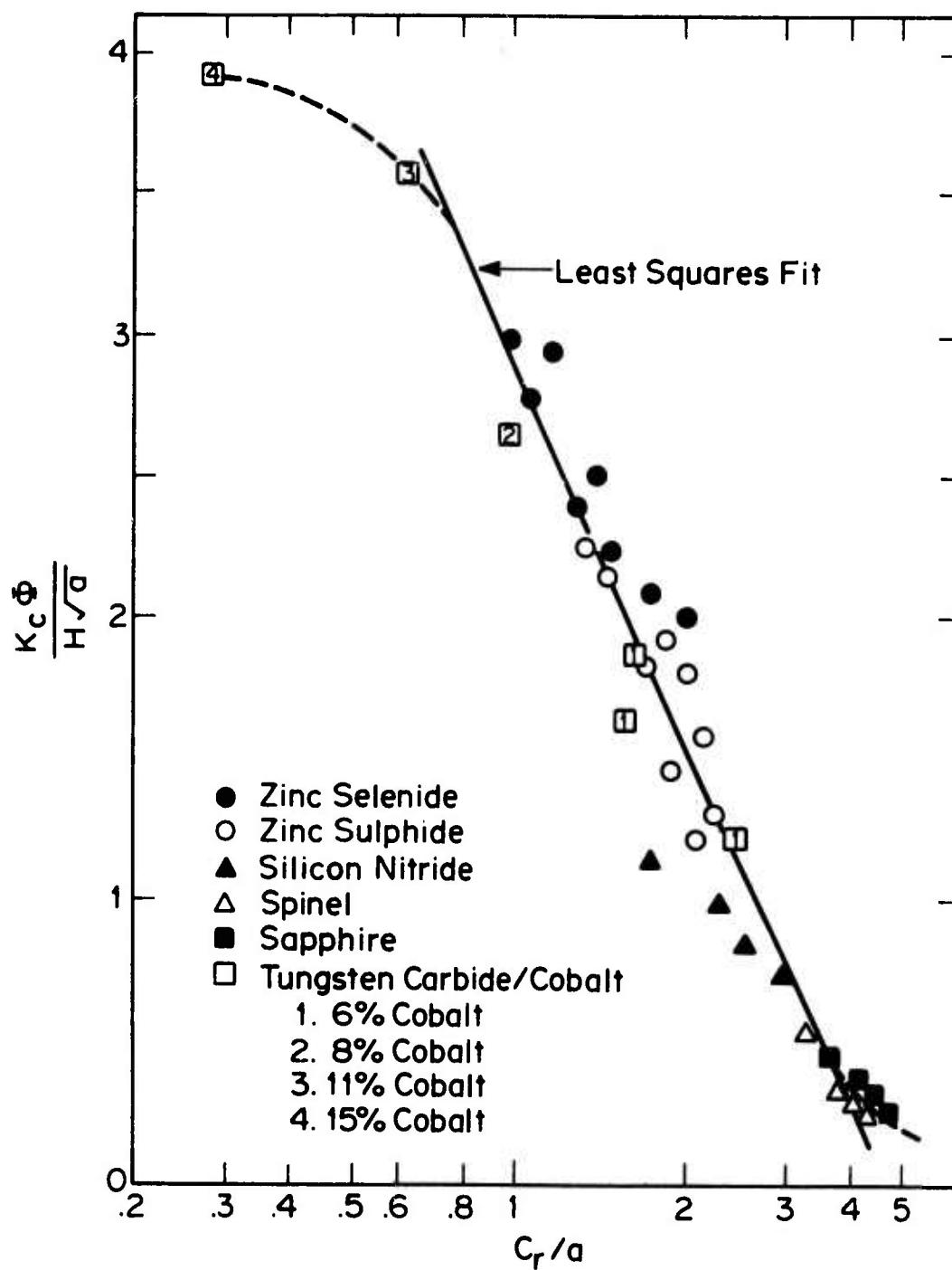


Figure 7. The relation between the logarithm of the normalized radial crack length C_r/a and $K_C \Phi / H \sqrt{a}$ for a range of brittle materials.

$$\frac{K_C \phi}{H\sqrt{a}} = 4.6 \log \left(\frac{4.5a}{C_r} \right) \quad (22)$$

where H is the hardness and ϕ is the constraint factor. Applying the hardness relation, for a fixed profile indenter,

$$H = \left(\frac{P}{2a^2} \right) \sin \theta, \quad (23)$$

where P is the applied force and θ is the semi-angle of the indenter, enables H to be eliminated, giving;

$$K_C = \left(\frac{2.3 \sin \theta}{\phi} \right) \frac{P}{a^{3/2}} \log \left(\frac{4.5a}{C_r} \right) \quad (24)$$

Since the parameters P , C_r and a are all directly measurable quantities, the possibilities for using indentation to measure fracture toughness become apparent. The pertinent calibration curve for a standard Vicker's indenter derived from Eq. (24), is plotted in Fig. 8.

The utility of indentation as a technique for approximate fracture toughness determinations lies in its simplicity and minimal material usage. The only test requirements are a hardness tester, and the only specimen requirements are a finely polished surface.* It may well prove, therefore, to be a useful routine test (of the quality control type) for evaluating fracture toughness. Also, the localization of the

*A good surface quality is needed to trace the radial crack extension, especially in opaque materials; care should also be taken, however, to avoid excess surface deformation during polishing because this can introduce compressive surface stresses and hence may result in erroneous toughness measurements.

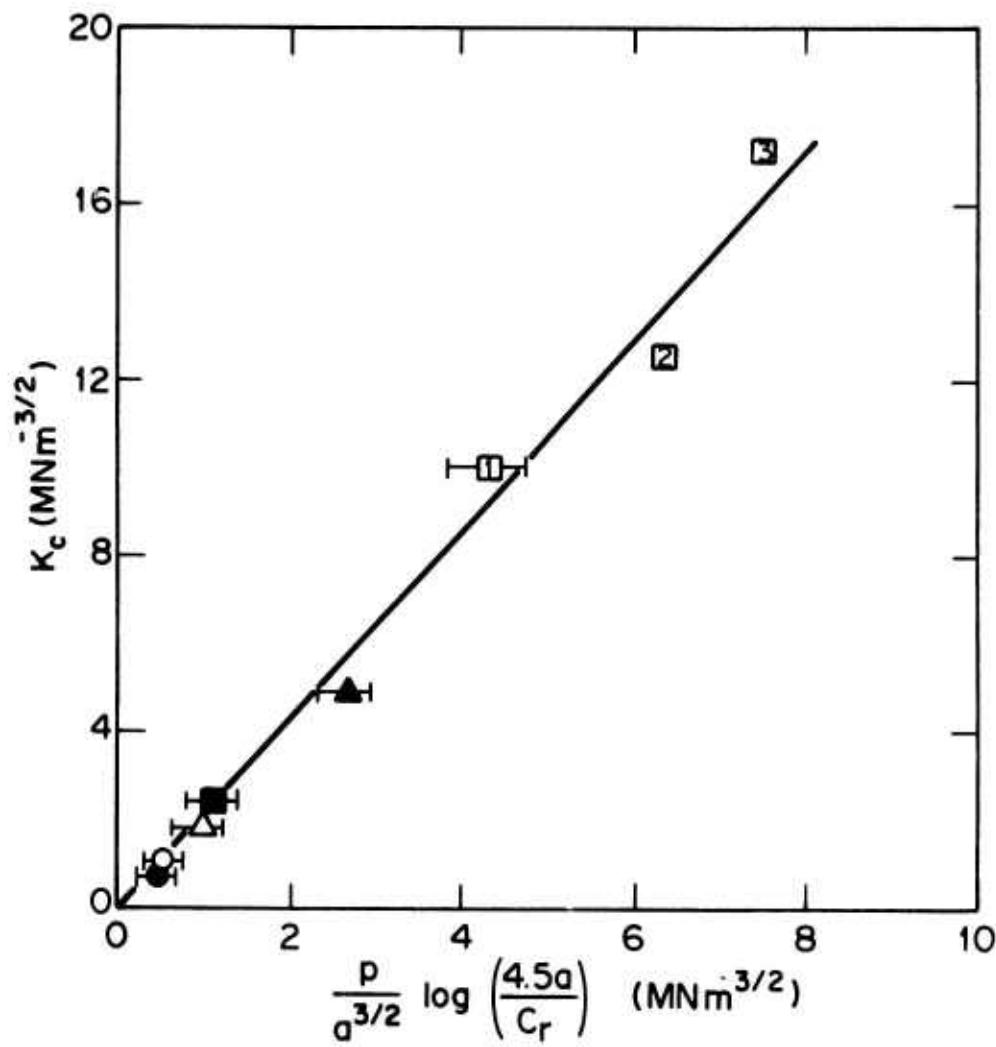


Figure 8. A calibration curve for fracture toughness determination by Vicker's indentation. Refer to Fig. 7 for symbol identification.

fracture enables toughness determinations to be performed on small microstructural features, such as single grains or grain boundaries, thereby permitting access to basic fracture information (which has not previously been amenable to experimental determination).

Additional testing is clearly needed to determine the ultimate merit of the technique; especially an examination of potential sources of variability (such as the effect of the indenter) and its extension to other materials.

B. Material Removal Estimates

Observations of lateral fracture have indicated that such cracks are potentially a major source of material removal during plastic indentation. The characterization of lateral fracture should thus provide a basis for estimating material removal rates in situations which involve plastic indentation. These should include many of the sliding indenter processes (such as abrasive wear, machining and sawing) and particle impact processes (such as solid particle erosion).

1. Abrasive wear or machining

The lateral crack extension is much the same in sliding indentation as in static indentation², and thus the single indenter characterization should suffice for first estimates of material removal parameters. Material removal by lateral fracture is assumed to occur when the lateral cracks from adjacent indenters intersect. For the sliding indenter case, therefore, this mode of material removal

will occur when an indenting particle is separated from a pre-existing linear indent by an amount $\leq r_o^c$. An upper limit for the volume of material removed by each particle \hat{V}_i is thus;

$$\hat{V}_i = 2r_o^c h_i l_i \quad (25)$$

where h_i is the depth of the lateral crack below the surface and l_i is the distance of sliding. The lateral crack depth, h , is a variable quantity, but its average value is roughly proportional to the impression radius²; in general

$$h = \omega a^\lambda \quad (26)$$

where ω and λ are constants ≈ 1 . Substituting for h (Eq. 26) and for r_o^c (Eq. 6) in Eq. (25) gives;

$$\hat{V}_i = 2.6 \omega \left(\frac{\sigma_y}{K_c} \right)^{2/3} a^{(4+3\lambda)/3} l_i \quad (27)$$

For abrasive processes the abrasive particles are generally angular indenters. The impression radius can thus be related to the force on the particle by Eq. (23), and the uniaxial yield strength can be related to the hardness through the constraint factor ($H = \Phi \sigma_y$). Hence, Eq. (27) becomes;

$$\hat{V}_i = 2.6 \left[\left(\frac{\sin \theta}{2} \right)^{(4+3\lambda)/6} \frac{\omega}{\Phi^{2/3}} \right] \frac{P_i^{(4+3\lambda)/6} l_i}{K_c^{2/3} H^{\lambda/2}} \quad (28)$$

where P_i is the vertical force on the particle. The upper limit for the volume removed by N abrasive particles is thus;

$$\hat{V} = \frac{\Lambda'}{K_c^{2/3} H^{1/2}} \sum_{i=1}^{i=N} P_i^{(4+3\lambda)/6} \ell_i$$

$$= \frac{\Lambda'}{K_c^{2/3} H^{1/2}} N \bar{P}^{(4+3\lambda)/6} \bar{\ell} \quad (29)$$

where $\Lambda' = 2.6 (\sin\theta/2)^{(4+3\lambda)/6} \omega_F^{-2/3}$. The important features of the predicted removal rate which we wish to emphasize are, the inverse dependence on the hardness and the fracture toughness and the direct dependence on the load. Taking λ as unity to obtain approximate numerical values for the exponents, gives the approximate proportionality;

$$\hat{V} \propto \frac{P^{7/6} \ell}{K_c^{2/3} H^{1/2}} \quad (30)$$

There are very few experimental measurements of abrasive material removal, under controlled test conditions, which can be compared with this predicted dependence. But, the data that are available are consistent with the prime dependence on the load, the hardness and the fracture toughness. The load dependence determined by Wilshaw and Hartley¹⁰ (Fig. 9) has an exponent ranging from 1.07 to 1.3; this range includes the exponent predicted by Eq. (30). The dependence on hardness and toughness can be determined from the data of Rice and Speronello,¹¹ because independent measurements of H and K_c are available for these materials (Table I). A plot of the reciprocal of the sawing rate (normalized to a grain size of $0.4 \mu\text{m}$)

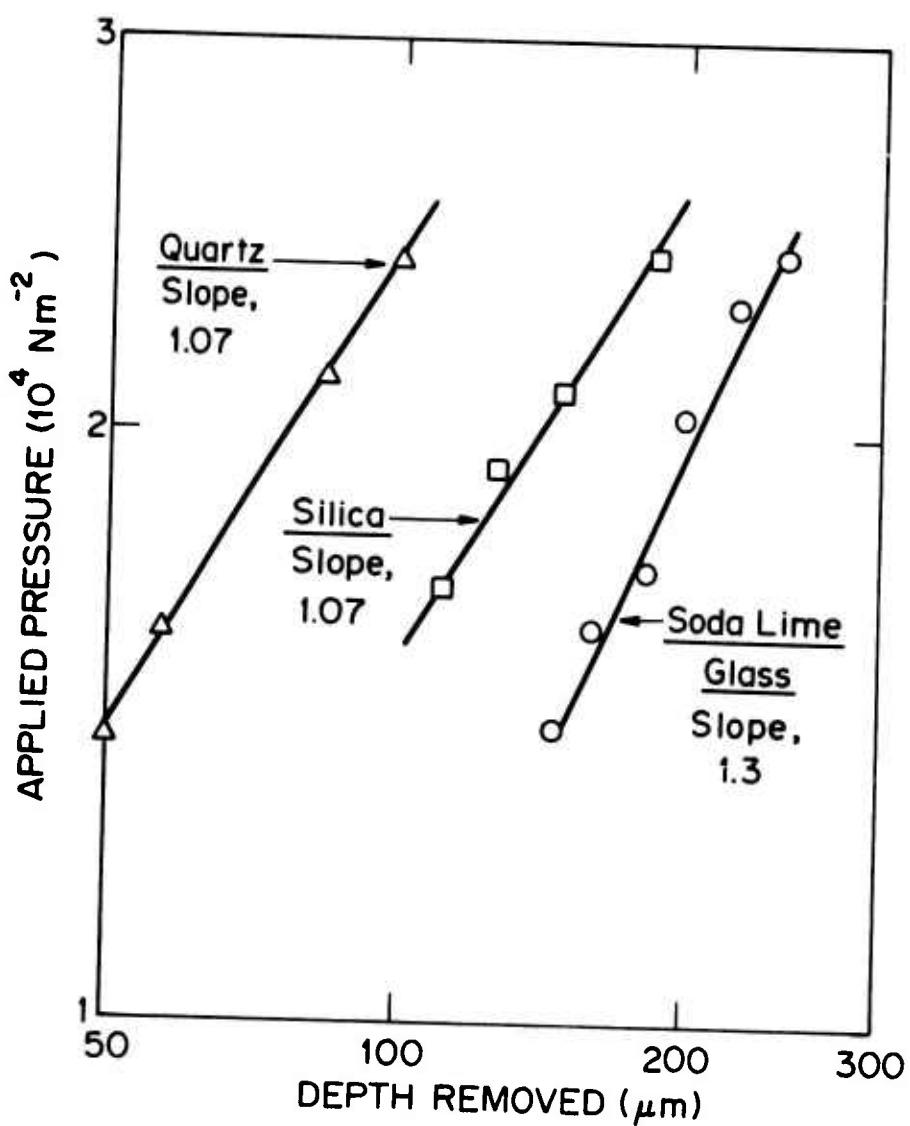


Figure 9. A logarithmic plot of the wear rate as a function of the applied load for three materials (data from Wilshaw and Hartley¹⁰).

TABLE I

Hardness and Fracture Toughness Data for the
Materials Used in the Sawing Study by Rice and Speronello¹¹

Material	K_C (MN m ^{-3/2})	H (GN m ⁻²)
B ₄ C	~6 ¹⁴	50 ¹¹
Si ₃ N ₄	5 ¹⁵	18 ¹⁶
Al ₂ O ₃	~4 ^{18, 14}	24 ¹¹
ZrO ₂	2.6 ¹⁴	15 ¹¹
MgO	~2.2 ¹⁷	7 ¹¹
MgO·Al ₂ O ₃	1.7 ¹⁷	13 ¹¹

versus the product $K_C^{2/3}H^{1/2}$ is shown in Fig. 10. The sawing rate variations are consistent with the predicted dependence on the hardness and the toughness. However, more detailed correlations should take account of possible frictional variations (because these data were obtained using a constant tangential force); also, the grain size dependence of the sawing rate should be effectively incorporated in the material removal analysis.*

The initial correlations of material removal by lateral fracture with observed abrasive removal rates are encouraging, and suggest that more detailed lateral fracture studies would prove invaluable for the analysis and prediction of abrasive material removal. Of specific interest would be a more extensive study of lateral fracture morphology, especially for interacting lateral cracks, to obtain more definitive exponents for each of the important material removal parameters.

2. Low velocity solid particle erosion

For solid particle erosion, material removal by lateral fracture is again considered to occur when the lateral cracks from adjacent impacts interact. An upper limit for the material removed by each particle can thus be obtained from r_o^C as

*Tentatively, we suggest that the grain size dependence is related to the relative change in the extent of the lateral fracture. For coarse grain sizes the plastic zone and the lateral cracks are contained with single grains, such that the single crystal H and K_C values pertain; while for fine grained material the plastic zone and the cracks extend over several grains so that the polycrystalline K_C and H values apply. A gradual transition in the effective K_C and H thus occurs as the grain size changes.

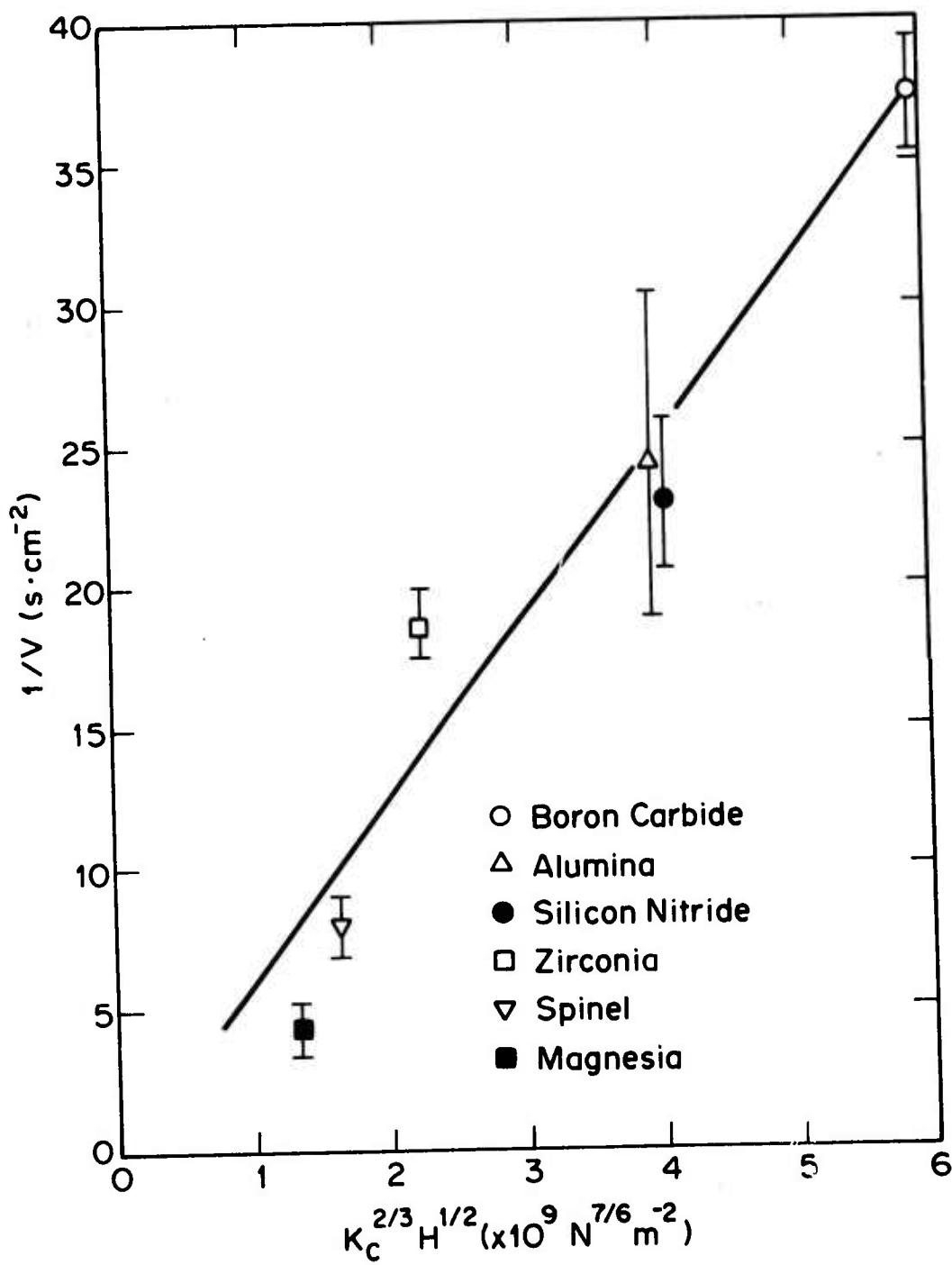


Figure 10. A plot of the reciprocal of the sawing rate versus the product $K_c^{2/3} H^{1/2}$ normalized to a grain size of $\sim 0.4 \mu\text{m}$ (from Rice and Speronello¹¹); the error bars indicate the typical sawing rate variations obtained for each material.

$$\hat{V}_j = \pi (r_o^c)^2 h_j \quad (31)$$

Substituting for r_o^c from Eq. (6) and for h_j from Eq. (26) gives;

$$\hat{V}_j = 1.7\pi \left(\frac{\sigma_Y}{K_D}\right)^{4/3} a^{(8+3\lambda)/3} \quad (32)$$

where K_D is the dynamic fracture toughness. For angular projectiles, a can be eliminated using Eq. (23) and σ_Y can be replaced by H_D/ϕ^{\dagger} (see section 1) to obtain;

$$\hat{V}_j = \frac{1.7\pi}{\phi^{4/3}} \left(\frac{\sin\theta}{2}\right)^{(8+3\lambda)/6} \frac{P_j}{K_D^{4/3} H_D^{\lambda/2}}^{(8+3\lambda)/6} \quad (33)$$

The upper limit for the total amount of material removed in unit time is thus

$$\hat{V} = \frac{\Lambda'' N P^{(8+3\lambda)/6}}{K_D^{4/3} H_D^{\lambda/2}} \quad (34)$$

where N is the number of projectiles that impact the surface in unit time and $\Lambda'' = 1.7\pi (\sin\theta/2)^{(8+3\lambda)/6} \phi^{-4/3}$. Further progress now requires P to be replaced by its functional dependence on the velocity, v_i . Unfortunately, the force cannot be described by a single function over the entire velocity range; however, a reasonable description can be provided by two functions (section III).

[†] H_D is the dynamic hardness, which is expected to be a function of the projectile velocity for most brittle materials.¹⁹

At low velocities, in the quasi-elastic regime defined by;

$$\left(\frac{v_i}{c_s^p}\right)^2 \left(\frac{G^p}{G^t}\right) \leq 4.5 \times 10^{-3} \left(\frac{G^t}{G^p}\right)^5$$

the force, velocity relation is given directly by Eq. (16).

Substituting this value for P in Eq. (34) gives;

$$\hat{V} = \frac{4.6^{(8+3\lambda)/6}}{K_D^{4/3} H_D^{\lambda/2}} \frac{\Lambda''}{N} \left((G^t)^2 v_i^6 b^{10} \rho^3 \right)^{(8+3\lambda)/30} \quad (35)$$

where $\rho (=G^p/(c_s^p)^2)$ is the density of the projectile. Approximate numerical estimates of the exponents are again obtained by setting $\lambda = 1$, giving;

$$\hat{V} \approx \frac{16\Lambda'' N (G^t)^{11/15} v_i^{11/5} b^{11/3} \rho^{11/10}}{K_D^{4/3} H_D^{1/2}} \quad (36)$$

The shear modulus is thus an additional material parameter (compared to the abrasive wear problem) affecting material removal, entering through its effect on the impact force. The fracture toughness again appears as a central parameter. The hardness is also important, but the hardness effect may be largely offset by the shear modulus, which tends to scale directly with the hardness in many brittle materials.²

At high velocities, in the elastic/plastic range, the force function is more complex (Eq. 15) and cannot be expressed in terms of the material and projectile parameters as a series

of power functions, except over a very limited velocity range. The principal material removal effects within this range are illustrated by examining a reduced form of Eq. (15) which applies for $P \gtrsim 10^5 P^*$;

$$P \approx 20 \frac{b^2}{(Gt)^2} \sigma_Y^3 \left[143 \left(\frac{v_i}{c_s^p} \right)^2 \left(\frac{G^p}{Gt} \right) \left(\frac{Gt}{\sigma_Y} \right)^2 \right] \quad (37)$$

which yields a material removal rate (for $\lambda = 1$) of;

$$\hat{V} \propto \frac{N}{K_D^{4/3}} \frac{H_D^{5/3}}{(Gt)^{11/3}} b^{11/3} \ln^{11/2} \left[143 \phi^2 v_i^2 \rho \frac{Gt}{H_D^2} \right] \quad (38)$$

The dependence of the material removal rate on the fracture toughness and the projectile size are essentially the same as in the quasi-elastic range. But the velocity exponent is a function of both the velocity itself, the hardness of the target, etc., such that the exponent decreases from its quasi-elastic value as the velocity or the hardness decrease. Similarly, the dependence of \hat{V} on hardness is a function of the velocity and hardness, with the role of hardness tending to diminish (compared to the quasi-elastic regime) as the velocity increases or the hardness decreases.

One important aspect of low velocity solid particle erosion by lateral fracture that has not been incorporated in our analysis is projectile fragmentation.¹² This process effectively reduces the input values of b and v_i and increases N , resulting in a lower dependence of \hat{V} on v_i and b . Also,

since fragmentation is more likely to occur with harder target materials, the fragmentation process can introduce an additional hardness dependence of the removal rate.

There are few solid particle erosion data in the literature that can be used to assess the essential features of the predicted erosion rate. Data by Oh et al,¹ Adler^{1,3} and Tilly and Sage^{1,2} give velocity exponents per particle of 2.5-3, 2.7* and 2[†], respectively, compared with a predicted value of 2.2 in the quasi-elastic range and <2.2 at higher velocities. Oh et al¹ also give a size exponent ranging from 3.2 to 5.2 (3.7 average) and Tilly and Sage^{1,2} give 2.2[†] compared to a predicted value of 3.7. These correlations are good, but they are not an effective test of the model, because other material removal mechanisms are expected to give similar velocity and size exponents.¹ A more critical test of the model would be a comparison of the dependence of erosion on dynamic fracture toughness or dynamic hardness. The only comparative data for brittle materials are Tilly and Sage's data^{1,2} for glass and silicon nitride. The erosion rates for these two materials differ by a factor of 200, which could not be explained by the relatively smaller

*The data by Adler^{1,3} and by Tilly and Sage^{1,2} were obtained by a whirling arm technique where N is proportional to v ; their exponents must therefore be reduced by ~ 1 to give the material removal per particle.

†Fragmentation occurs in Tilly and Sage's experiments, which accounts in part for the lower values of their velocity and size exponents.

difference (a factor of ~13) in toughness, hardness, modulus parameter suggested by Eq. (36), (although the prediction is, of course, qualitatively correct). Extensive fragmentation occurs in their experiments, however, and this may substantially contribute to the low erosion rates for silicon nitride, as noted above.

V. CONCLUSIONS

1. The characterization of radial fracture from plastic indentations has been shown to provide a good measure of the fracture toughness of a wide range of brittle materials.
2. An evaluation of lateral fracture as a prime source of material removal by plastic indentation processes, such as abrasion or solid particle impact, has yielded encouraging correlations with available material removal data for brittle materials.

ACKNOWLEDGEMENT

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

APPENDIX I

Analogous Stress Field for Lateral Fracture

Approximate analytic solutions for lateral fracture problems can be obtained by commencing with the stress intensity factor relation for a penny-shaped crack in an axisymmetric prior stress field⁶;

$$K = \frac{2}{\sqrt{\pi C}} \int_{C_0}^C \frac{\sigma r}{\sqrt{C^2 - r^2}} dr \quad (A1)$$

where r is the distance from the center of the impression and C_0 is the position of the inner crack front. Usually, the prior stress σ is known and K is computed. However, in this instance, K_C is known (Appendix II);

$$K_C = \Lambda^{4/3} \sigma_Y \frac{a^{11/6}}{C^{4/3}}$$

and we require to solve Eq. (A1) for σ . Since lateral crack propagation is stable in the regime of interest, K can be set equal to K_C to obtain the prior stress over the crack trajectory, giving;

$$\int_{C_0}^C \frac{\sigma r}{\sqrt{C^2 - r^2}} dr = \frac{\sqrt{\pi}}{2} \sigma_Y \Lambda^{4/3} \frac{a^{11/6}}{C^{5/6}} \quad (A2)$$

If we now assume that σ over the crack path has the form,

$$\sigma = \chi r^{-n} \quad (A3)$$

where χ and n are constants, integration of the left hand side of Eq. (A2), signified by J_L , gives;

$$J_L = \frac{\chi \sqrt{1 - (C_0/C)^2}}{(n-2)CC_0^{n-2}} + \left(\frac{n-3}{n-2}\right) \frac{\chi}{C^2} \int_{C_0}^C \frac{r^{3-n}}{\sqrt{C^2-r^2}} dr \quad (A4)$$

Expanding the square root in the first term on the right hand side of Eq. (A4) then gives;

$$J_L = \frac{1}{C} \left[\frac{\chi}{(n-2)C_0^{n-2}} \right] \left[1 - \left(\frac{C_0}{C} \right)^2 - \frac{3}{4} \left(\frac{C_0}{C} \right)^4 - \dots \right] + \left(\frac{n-3}{n-2} \right) \frac{\chi}{C^2} \int_{C_0}^C \frac{r^{3-n}}{\sqrt{C^2-r^2}} dr \quad (A5)$$

Rewriting the right hand side of Eq. (A2), J_R , as;

$$J_R = \frac{\sqrt{\pi}}{2} \sigma_Y \Lambda^{4/3} \frac{a^2}{C} \left(\frac{C}{C_0} \right)^{1/6} \quad (A6)$$

and expanding the last term gives;

$$J_R = \frac{1}{C} \left[\frac{\sqrt{\pi}}{2} \sigma_Y \Lambda^{4/3} a^2 \right] \left[1 + \ln \frac{(C/a)}{6} + \frac{\ln^2(C/a)}{72} + \dots \right] \quad (A7)$$

Dividing Eqs. (A5) and (A7) gives;

$$\begin{aligned} \frac{J_L}{J_R} &= \frac{2\chi}{\sqrt{\pi}(n-2)\Lambda^{4/3}\sigma_Y a^2 C_0^{n-2}} \left\{ \frac{[1 - 1/2(C_0/C)^2 - 3/4(C_0/C)^4 - \dots]}{[1 + (1/6)\ln(C/a) + (1/72)\ln^2(C/a) + \dots]} \right\} \\ &+ \left(\frac{1}{C} \right) \frac{2\chi(n-3)}{\sqrt{\pi}(n-2)\Lambda^{4/3}\sigma_Y a^2 [1 + (1/6)\ln(C/a) + (1/72)\ln^2(C/a)]} \int_{C_0}^C \frac{r^{3-n}}{\sqrt{C^2-r^2}} dr \end{aligned} \quad (A8)$$

For $n = 3$ the last term on the right hand side is zero and for $3C_o \lesssim C \lesssim 4a$, the first term varies with C by $<25\%$. The quantities J_L and J_R are thus approximately equivalent, i.e., exhibit the same dependence on the common variable C , when;

$$n = 3$$

$$\chi = \frac{\sqrt{\pi}}{2} \Lambda^{4/3} \sigma_Y a^2 \quad (A9)$$

Substituting these values for n and χ into Eq. (A3) gives the stress field over the lateral crack trajectory as;

$$\sigma \approx \frac{\sqrt{\pi}}{2} \Lambda^{4/3} \sigma_Y a^2 C_o r^{-3} \quad (A10)$$

This result can be verified by substituting this value for σ into Eq. (A2) and integrating. Lateral fracture observations² indicate that the inner crack front, C_o , roughly coincides with the position of the plastic boundary, viz., $a < C_o \lesssim 1.3a$; the most stringent limitation on the application of Eq. (A10) is thus, $C/a \gtrsim 4$. Since most fractures of interest are within this range (see Fig. 12), and since the term, $\ln(C/a)/6$, does not vary rapidly for $C/a > 4$, Eq. (A10) can be considered suitable for approximate use in most lateral fracture analyses.

Rearranging Eq. (A10);

$$\frac{\sigma}{\sigma_Y} = \frac{\sqrt{\pi}}{2} \left(\Lambda^{4/3} \frac{C_o}{a} \right) \left(\frac{a}{r} \right)^3 \quad (A11)$$

indicates its resemblance to the stress field (outside the plastic zone) for a pressurized spherical cavity⁴;

$$\frac{\sigma}{\sigma_Y} = \frac{\xi^3}{3} \left(\frac{a}{r} \right)^3 \quad (A12)$$

where ξ is a dimensionless term that depends on the position of the plastic boundary and a is now the position of the inner surface of the cavity. Since the plastic boundary also determines the position of the inner crack front (see above), the analogy relates ξ to the inner crack front by:

$$\xi = 1.38 \Lambda^{4/9} \left(\frac{C_0}{a} \right)^{1/3} \quad (A13)$$

For the existence of an inner crack front, the stress intensity factor at C_0 must always be less than K_C . A method for determining K at the inner front is not available, but for the uniform tension case (and $C \approx 3C_0$), K at the inner front exceeds K at the outer front by⁶; $(\pi^{3/2}/2)(C/C_0)^{1/2}$. It is evident, therefore, that to prevent inward propagation the stress immediately beyond the inner crack front must be compressive and substantially larger in magnitude than the tensile stress, as shown schematically in Fig. 11. The stress field for a pressurized spherical cavity exhibits a similar tendency within the plastic zone ($x < \xi a$), as shown in Fig. 11. For the analogy to be consistent, therefore, we note from Fig. 11 that ξ must satisfy the condition, $\xi \approx C_0/a > 1$. Substituting the value of Λ from Fig. 12 (0.51) into Eq. (A13) shows that $\xi = 1.02 (C_0/a)$, and as indicated above, $C_0/a > 1$. The basic analogy thus appears reasonable.

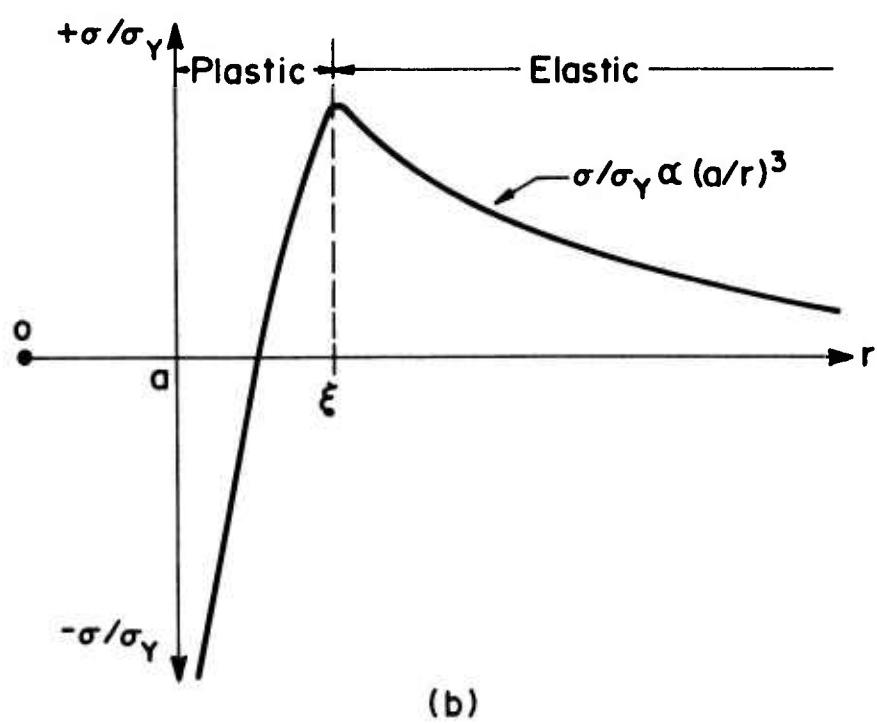
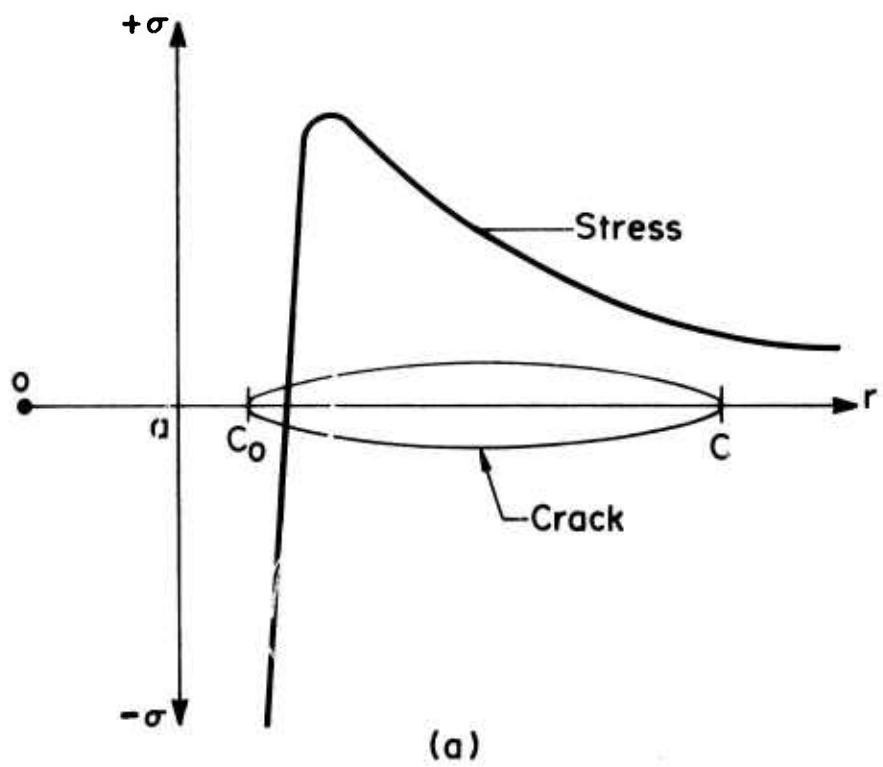


Figure 11. A schematic of (a) the stress field expected over the lateral crack trajectory, compared with (b) the stress field for a pressurized spherical cavity.

APPENDIX II

Lateral Crack Lengths

For material removal estimation it is convenient to characterize the position of the lateral crack front from the center of the impression, because the impression itself generally tends to be removed. In a previous paper², lateral crack lengths were determined from the impression periphery. These data are thus re-evaluated by adding the impression radius a to the lateral crack length to obtain the new length parameter C . The results are plotted in Fig. 12 using the dimensionless variables C/a and $K_c/\sigma_Y \sqrt{a}$. A linear fit to the data* then gives the relation;

$$C = \Lambda \left(\frac{\sigma_Y}{K_c} \right)^{0.74} a^{1.37} \approx \Lambda \left(\frac{\sigma_Y}{K_c} \right)^{3/4} a^{11/8} \quad (A14)$$

where Λ is a constant equal to 0.51.

*The linear fit is not necessarily considered to be the best fit to the data, but the fit is reasonable and it yields functional relations that are analytically tractable.

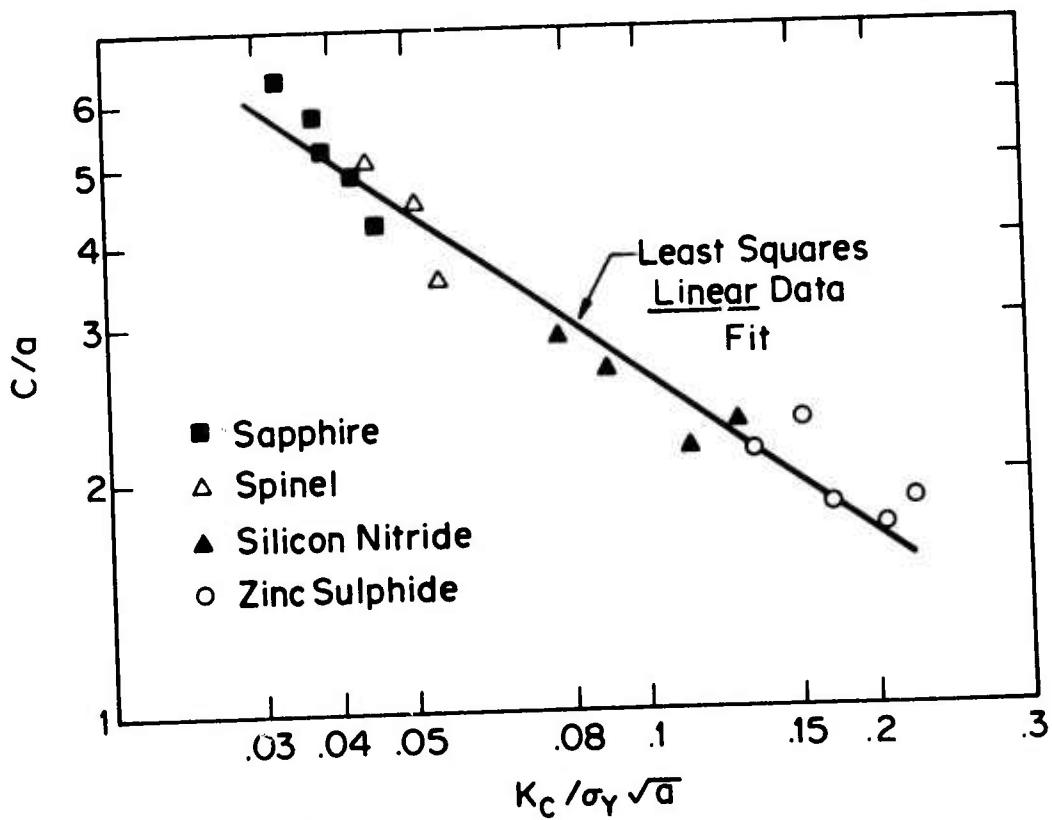


Figure 12. A logarithmic plot of the normalized lateral crack radius C/a as a function of the normalized stress intensity factor, $K_c/\sigma_Y \sqrt{a}$.

References

1. H. L. Oh, K. P. L. Oh, S. Vaidyanathan and I. Finnie, N.B.S., Special Publication 348 (1972) p. 119.
2. A. G. Evans and T. R. Wilshaw, "Quasi-Static Solid Particle Damage in Brittle Materials - I. Observation and Analysis," to be published.
3. B. R. Lawn and T. R. Wilshaw, J. Mat. Sci. 10 (1975) 1049.
4. R. Hill, Plasticity, Clarendon Press, Oxford (1950) p. 97.
5. A. G. Evans and T. G. Langdon, Progress in Materials Science, in press.
6. G. C. Sih, Handbook of Stress Intensity Factors, Lehigh University, Bethlehem, Pa. (1973).
7. S. Timoshenko and J. N. Goodier, Theory of Elasticity, McGraw-Hill, New York (1951) p. 383.
8. Y. M. Tsai, Int. J. Solids Structures 7, (1971) 543.
9. C. Hardy, C. N. Baronet and G. V. Tordion, Int. J. Num. Methods Eng. 3, (1971) 451.
10. T. R. Wilshaw and N. E. W. Hartley, Third European Symposium on Comminution (Verleg Chemie GMBH, Weinheim/Bergstrasse, 1972) p. 33.
11. R. W. Rice and B. K. Speronello, to be published.
12. G. P. Tilly and W. Sage, Wear 16 (1970) 447.
13. W. F. Adler, work performed at Bell Aerospace.
14. R. W. Rice, Fracture Mechanics of Ceramics (Ed. R. C. Bradt, F. F. Lange and D. P. H. Hasselman) Plenum, New York (1974) p. 323.
15. A. G. Evans and S. M. Wiederhorn, J. Mat. Sci. 9 (1974) 270.
16. D. J. Godfrey and K. C. Pitman, Ceramics for High Performance Applications, (Ed. J. J. Burke, A. E. Goram and R. M. Katz), Brook Hill, Mass. (1975) p. 425.

17. L. J. Graham and G. Alers, Characterisitcs of Acoustic Emission Signals From Ceramics, Science Center Report SC544.14FR, Feb. 1974.
18. L. A. Simpson, Fracture Mechanics of Ceramics, loc cit, p. 567.
19. T. R. Wilshaw and A. G. Evans, "Solid Particle Impact Damage in Brittle Materials," to be published.

CONTRIBUTION ON THE ISSUES OF WEAR
PARTICLE SIZE, COMPOSITION AND STRUCTURE

A. W. Ruff

Studies have been conducted on many samples of wear debris obtained from a variety of tests under both dry and lubricated conditions. Some of the tests were conducted under controlled laboratory conditions but many others involved the output from field-operating systems. The more significant and common observations are noted here including some brief comments on possible implications in connection with particular wear mechanisms. The wear debris was recovered from the oil lubricant (the final wash in dry tests) using the Ferrograph instrument. When properly used it is our experience that this technique is a highly reliable method of (iron) wear particle recovery. Scanning electron microscope examination of the particles has been conducted directly on the Ferrogram.

Particle Size

The particle size distribution on a Ferrogram from a severe (rubbing) wear system extends from the sub-micron level ($\sim 0.1\mu\text{m}$) to about $10\mu\text{m}$. A few particles may be present of size up to $50\mu\text{m}$ (rarely to $100\mu\text{m}$). The peak in size distribution may be about $5\mu\text{m}$. In a low wear situation, the particle

sizes are 1/2 to 1/5 as large and particles over $5\mu\text{m}$ are rare.

It seems likely that many of the sub-micron particles may result from larger particle fragmentation during formation or while being recirculated through the bearing area. Thus mechanisms for particle formation need to explain sizes in the range $1\mu\text{m}$ to $10\mu\text{m}$, and to $50\mu\text{m}$ in heavy wear situations.

Particle Shape

In moderate to severe wear situations the predominant shape is disc-like with the small dimension about 1/5 that of the other two (equal) dimensions. Sub-micron particle shapes are as likely to be equi-axed as disc-like (but these may not be virgin particles). Large particle shapes ($>10\mu\text{m}$) vary more and fewer have been observed, however, they too generally have a thin dimension of the order of $2\mu\text{m}$ to $10\mu\text{m}$. In low wear situations the predominant shape tends more away from disc-like toward an equiaxed configuration. Naturally all particle collections involve a large spread in size and shape; the above comments are meant to describe the prominent situations. Quantitative size and shape distributions of good statistical significance are not available and need to be done.

Particle Composition

Electron probe X-ray emission experiments have been conducted on many individual particles. Calculations have been done on the size effect for X-ray emission and corrections can be made for the size effect (in iron) below $3\mu\text{m}$. The prominent wear debris in the size range $1\mu\text{m}$ to $15\mu\text{m}$ invariably

shows the same X-ray emission as that obtained from the bulk (bearing) surface except that a small reduction in $Fe-K_{\alpha}$ line intensity is usually found (about 10 to 20% less). X-ray powder diffraction studies have been conducted on collections of wear particles. Several types of iron oxide are usually found, depending on the test conditions. Generally $\gamma-Fe_2O_3$ and FeO are identified, in some instances $\alpha-Fe_2O_3$, and Fe_3O_4 . The oxide lines are generally strong and indicate a significant relative amount present. The α -Fe diffraction lines are generally strong and quite broad (indicating considerable plastic deformation). Electron (transmission) diffraction studies have been conducted on perhaps 100 different particles and the results are in agreement with the X-ray diffraction data. The reduced $Fe-K_{\alpha}$ emission from the wear particles is a result of the oxide content on and within the particles.

Particle Structure

The only detailed information available in our work on the internal structure of the wear particles results from transmission electron microscope studies of individual particles removed from various Ferrograms. [Electron channeling studies were attempted on many particles without success, indicating plastic strains in excess of about 20% or possibly surface coatings of oxide or strained metal that destroyed the channeling effect.] In nearly all cases, it was not possible to image internal details (e.g., sub-boundaries, dislocations) in the particles at 200kV electron beam condi-

tions and penetrations of about $0.5\mu\text{m}$ to $0.8\mu\text{m}$. In a few cases sub-boundary structures were seen throughout the particle, however, it is believed that those particles had recovered or recrystallized prior to observation. Diffraction patterns were broad and diffuse in character indicating that considerable plastic deformation remained in the metal. Some particles were found to be principally oxide (50 to 90%) however, these were generally thin overall ($\approx 0.5\mu\text{m}$) and not representative of the predominant type. The latter were found to have an oxide layer overall, perhaps 300\AA to 1000\AA thick surrounding a deformed metal core. The thicker regions of these particles could not be penetrated or studied with any resolution. Clustering of sub-micron oxide particles was frequently observed, so that a larger appearing particle on close study would prove to be a porous, composite material.

In summary, it must be recalled that the age and experience associated with each wear debris particle subsequent to separation from the wearing surfaces is largely unknown. Hence, considerable variability in analytical findings is not surprising. It is probably necessary to conduct additional, carefully controlled wear experiments involving prompt collection and analysis of debris particles in order to reduce the observed variability and provide more quantitative and accurate data. The role of the oxide (and contaminants) at the wearing surfaces is believed important and needs further study.

COMPARISON BETWEEN THE ADHESIVE AND FATIGUE
EXPLANATIONS OF "ADHESIVE WEAR"

Ernest Rabinowicz

As has been made clear at the meeting, there are two completely different explanations for the phenomenon known as adhesive wear, (adhesive wear being defined as material removed from surfaces during sliding, in the absence of special factors such as undue abrasiveness, corrosion or brittleness of the surfaces and their environment).

According to the classical theory, which probably dates back to Bowden, Moore and Tabor in 1943, there is strong adhesion between some of the junctions at the interface, and adhering wear particles are ripped out. Eventually, for reasons discussed by Rabinowicz in 1961, these often become loose wear particles.

The fatigue explanation, which probably was first propounded by a number of workers at IBM (Ku, Brophy, etc.) in the late 1950's, and has recently been revived and extended by Suh, starting in about 1971, suggests that continual sliding over a surface weakens it, eventually cracks form and grow, and wear particles are produced.

The two theories make different predictions about the sizes and shapes of wear particles. Leaving out of account

any auxiliary hypotheses, the adhesion theory predicts that for unlubricated sliding of metals, the wear particles will be of diameter between 10 and 100 microns and equi-axed, while in the presence of lubricants they will be a factor of ten smaller (1 to 10 microns) but of the same shape. According to the fatigue theory, the particles will be flat sheets of thickness about 1 micron and of poorly specified length and width.

What is the experimental evidence? The adherents of the adhesion theory can point to a couple of studies (by Rabinowicz and by Finkin) which suggest that wear particles are indeed chunky, with typical relative dimensions $3 \times 2 \times 1$. Supporters of the fatigue theory can find evidence of many thin flat particles, with typical dimensions $15 \times 10 \times 1$, and moreover have produced sections of worn surfaces showing long cracks just below the surface, running parallel to the surface, and obviously the site of a new potential wear particle. (At this meeting, V. Westcott showed many illustrations of both types of particles.)

How can these contradictory views be reconciled. First, it should be emphasized that in a typical wear experiment a vast number of wear particles may be produced, covering a wide range of sizes (from $10^{-3} \mu\text{m}$ to $10^2 \mu\text{m}$ in a typical case) and shapes (from perfect sphere to long thin ribbon). The real question is, what does the typical, characteristic wear particle look like?

The adhesive wear proponents are interested in the question, how much material is worn off the surfaces, and what does this material consist of. Typically, about 95% of all the matter removed from a surface is contained in those 0.01% of all the wear particles with dimensions between 10 and 100 μ m. These particles are equiaxed, and their presence suggests that adhesive processes are indeed the main reason for loss of material during sliding. At the same time, the fact that the overall Archard wear coefficient is 10^{-4} suggests that these wear particles form at only 0.01% of the junctions, and that at the other junctions other things are happening.

The proponents of fatigue explanations look at surfaces, section surfaces, and look at wear particles, and find everywhere evidence of cracks below the surface, and flake-like particles suggesting that flake formation is the predominant mechanism of wear.

It is the purpose of this comment to suggest that the two points of view are not really different, that the adhesive model refers to the one-in-ten-thousand junctions where really severe interaction is occurring, while the fatigue model describes what is happening at the other 9999 out of 10000 junctions.

What is in dispute is which of these events is more important. Perhaps each controls the wear rate in different situations, the adhesive model when k is high ($k > 10^{-4}$), the fatigue model when k is low ($k < 10^{-4}$).

My feeling is that members of the Materials Research Council have done a fine job of exposing the conflicts and the areas of agreement between the two types of theories. What is needed at this point is:

1. Careful experiments in which wear particles are produced in various ways and then carefully studied, to determine the significance of the various types of wear. This will make it clear what kinds of changes will reduce the wear rate.
2. Further work on the fatigue model to bring it to the point where it can make predictions about the wear coefficient which can be tested.
3. Further work on the adhesive wear model, to bring it to the point where it can make predictions about the wear coefficient which can be tested.

ION IMPLANTATION AND WEAR

J. W. Butler

ABSTRACT

Ion implantation has been used in a large number of applications to change the physical and chemical properties of surfaces. But ion implantation has only recently been applied to studies of friction and wear. Ion beams of Kr, Sn, In, Ag, Pb, Mo and S have been implanted into steel at energies between 100 and 400 KeV and at fluences between 10^{16} and 10^{17} ions/cm². Measurements of friction between a tungsten carbide ball and the resulting implanted surfaces in air have been made by Hartley, Dearnaley and Turner, and these results compared with the unimplanted surfaces. The Sn implanted surface showed a decrease in friction by a factor of 2.7, the Pb implanted surface showed an increase in friction of a factor of 1.4, and the other implanted surfaces showed minor changes or no change. Hartley has made pin-and-disk measurements of wear of steel surfaces implanted with Mo and N ions and found substantially less wear of the pin when the disk was implanted than for the unimplanted disk. These pioneering experiments are not definitive, and further experiments are needed before a definitive conclusion can be drawn as to the effectiveness of ion implantation in reducing friction and wear of surfaces.

ION IMPLANTATION AND WEAR

J. W. Butler

INTRODUCTION

Ion implantation has been successfully used for a decade or more to modify the physical or chemical properties of surfaces for a number of different purposes; but only during the past three or four years has ion implantation been used in studies of friction and wear of surfaces. The present paper is a description of ion implantation as a technique, its advantages and disadvantages, and a discussion of some preliminary results of experiments on the friction and wear of ion implanted surfaces.

WHAT IS ION IMPLANTATION?

Ion implantation involves ionizing atoms of the desired species, accelerating these ions to relatively high speeds, and then driving them into the desired substrate, for the purpose of modifying the physical or chemical properties of the host material. Ion speeds usually lie within the range 10^{-2} to 1% of the speed of light; and deposition depths are usually in the range 10^2 \AA to 10^4 \AA . Relative concentrations of the implanted ions are usually in the range 0.5% to 5%.

The modified properties might be electrical, optical, chemical or mechanical. For example, ion implantation may be used to modify the behavior of a superconducting material, a

material's index of refraction, its charge-carrier concentration, its corrosion resistance, or its surface hardness. The solid may be crystalline, polycrystalline or amorphous.

DEPTH OF IMPLANTATION

When a fast atomic ion enters a solid material, the incident ion collides with the atoms of the material with four principal results. (1) Electrons are knocked out of the struck atoms, (2) Electrons are usually knocked out of the incident atomic ion, (3) The target atoms are usually displaced from their lattice sites, and (4) The incident ion loses energy and slows down.

The fourth effect limits the range of directly implanted ions to about one micrometer or about 10^4 Angstroms. The energy-loss process is complicated because of the number of parameters that are involved and the number of factors that affect the value of each parameter. For example, what is the average charge on the incident ion? And which electrons in the incident and target atoms are most directly involved in the collision process? Another complication is the statistical nature of the energy-loss process. Because of the statistical nature of the energy-loss process, not all identical ions that enter a solid with the same energy stop at the same depth. Hence ions implanted at the same energy occupy a substantial region of depth.

In spite of all the complications involved in the energy-loss process, a reasonably accurate mathematical theory

of the process does exist; and one can use this theory to calculate the depth distribution of implanted atoms.

RADIATION DAMAGE

One of the major disadvantages of ion implantation as a process in the fabrication of some finished materials is the radiation damage produced in the host material. The source of this radiation damage is the collisions between incident ion and target atoms referred to earlier. For most materials the damage is almost completely the result of displaced atoms or crystal defects or lattice disorder.

The radiation-damage problem is usually solved by some sort of annealing process. The lattice imperfections consist of vacant sites and interstitial atoms (sometimes called fresnel pairs). At a sufficiently high temperature (which depends on the nature of the host atoms and of the impurity atoms) these vacancies and interstitial atoms migrate and re-constitute the appropriate lattice. This repair of the radiation damage is an essential part of the application of ion implantation to some practical problems. However, in the experiments that have been performed so far in the areas of friction and wear, the radiation damage was not annealed out; and perhaps the stresses associated with the radiation damage are desirable.

EQUIPMENT

Ion implantation accelerators are somewhat bulky, complicated, and expensive; however, they are no more bulky

complicated, or expensive than many other items in a modern laboratory or factory.

Essentially an ion implantation accelerator consists of an ion source (which provides the desired atomic species in a form that can be accelerated), a high voltage source, an evacuated acceleration tube, and a target chamber containing the surface to be implanted. Usually a magnet is used to separate the beam into isotopic and energy components.

Usually in research into the effects of ion implantation, one needs to make an independent measurement of the depth concentration profile of the implanted ions. Such a measurement is most accurately obtained by one of two methods: (1) Rutherford backscattering of the incident ions or (2) nuclear reactions.

The former method is particularly well suited to cases in which heavy atoms are implanted in a light-atom substrate (in which case billiard-ball mechanics prevails, and the incident ions that bounce off the heavier atoms have more energy than those that bounce off the lighter atoms, just as a golf ball bounces back faster from a bowling ball than from a billiard ball). The depth scale is obtained from the fact that fast charged particles lose energy as they traverse matter. The energies of the backscattered particles are measured either by a magnetic spectrometer or by a solid-state detector. Figure 1 illustrates the type of information that one gets from this technique. In this illustrative example, thin layers of

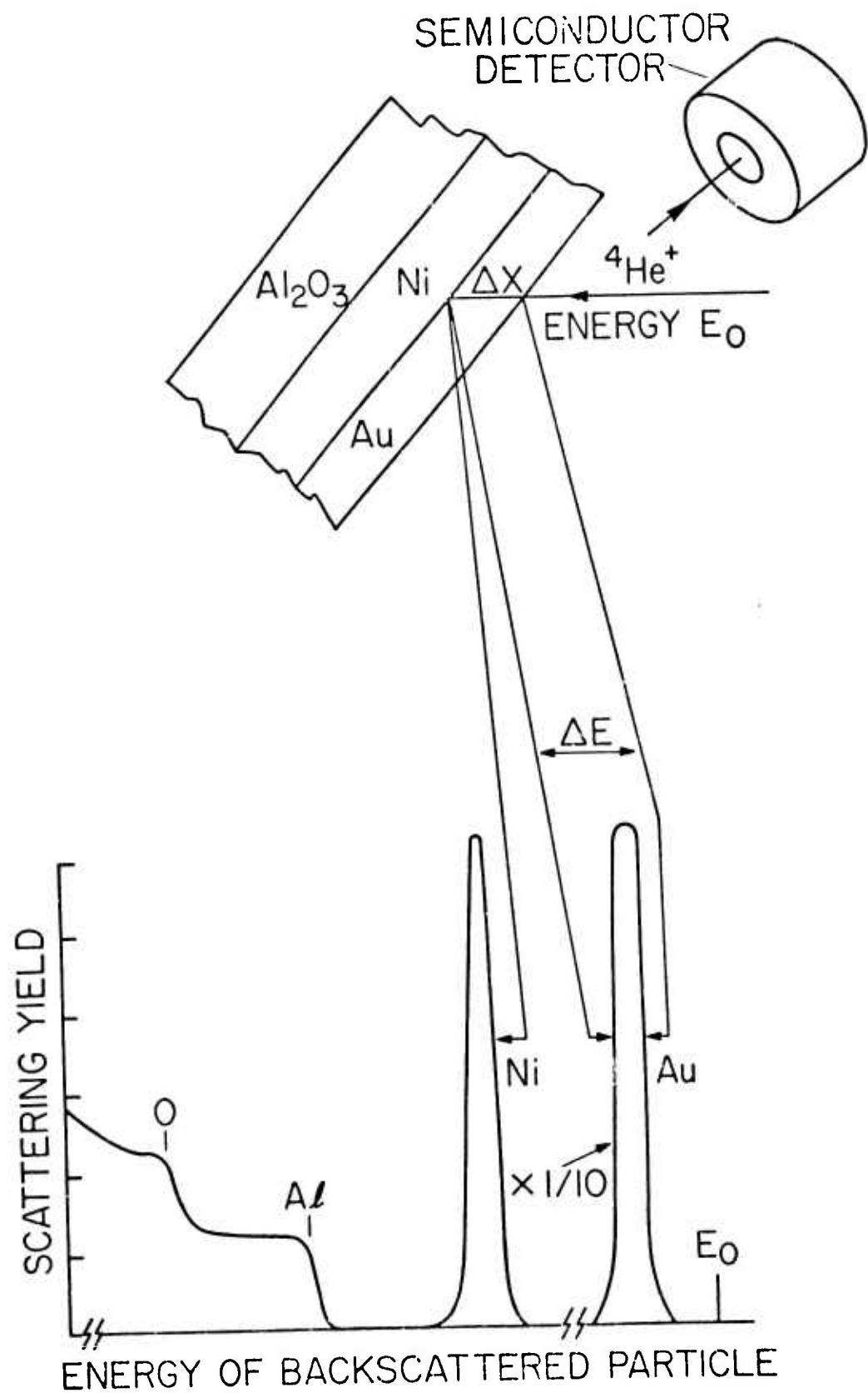


Figure 1. The analysis of a surface by Rutherford backscattering.

nickel and gold have been deposited on an Al_2O_3 substrate, and helium ions are used to probe the sample. The backscattered helium ions are detected and their energy measured in this example by a solid-state detector as illustrated here. Note how the scale of energy of the backscattered particle becomes a combination depth-and-mass scale for the atoms on the target. In the case of the thick Al_2O_3 substrate, one obtains a plateau for each mass rather than a peak because no matter how low one looks on the backscattered-particle energy scale, there corresponds a depth in the Al_2O_3 , where the helium ion can backscatter and re-emerge with that particular energy.

The second type of analysis of implanted layers -- nuclear reactions -- is performed with the aid of a precision analyzer which measures the energies of the incident particles producing the reaction. Figure 2 shows the sort of information one gets from this technique. This technique is based on the fact that the probability of a proton being captured by a given isotope and a gamma ray emitted is a very sensitive function of the energy of the proton. For some isotopes this probability, or "cross section" as it is usually called, is almost like a delta function or spike. This spike in the cross section is called a "resonance" by analogy with mechanical and electrical resonances. So in a sense the isotope serves as a sensor for resonance energy. Hence if a given material (say aluminum) is a certain depth beneath the surface of a substrate (say SiC), then for the proton energy to be at resonance at the depth of

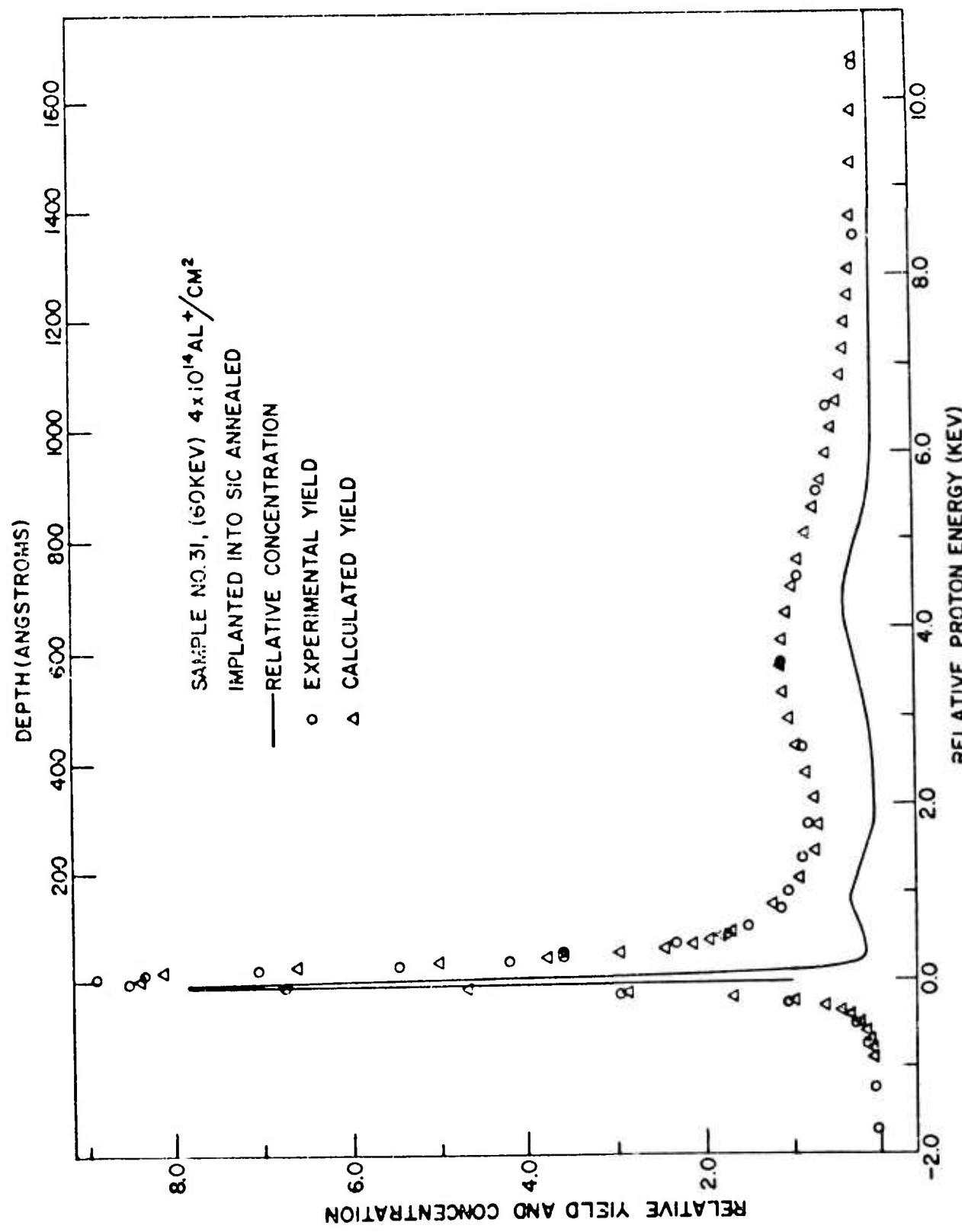


Figure 2. The analysis of a surface by nuclear reaction.

the aluminum, the bombarding energy must be above resonance energy by just the amount of energy the proton loses in reaching the depth of the aluminum. In Figure 2 the solid curve shows the actual distribution of aluminum in a SiC substrate, the open triangles show the calculated yield of gamma rays for this target as a function of bombarding energy above the resonance energy. And the open circles show the experimental yield of gamma rays. The agreement is excellent. In practice one works back the other way, of course. One starts with an experimental yield curve and deduces from it the distribution of aluminum. The calculated curve is the result of a sophisticated computer program that is a triple integration and takes into account such characteristics as the statistical fluctuations of energy loss of the protons.

EXPERIMENTS ON FRICTION

Since the application of ion implantation to the field of friction is so new, the nature of the effect is not understood, and empirical evidence is quite limited. Hartley at the University of Sussex and Dearnaley of the Atomic Energy Research Establishment at Harwell¹ implanted ions of soft "solid-lubricant-type" metals (e.g., tin, lead, and silver) into steel surfaces, and measured the coefficients of friction of the implanted and unimplanted surfaces. In one experiment they dragged a loaded tungsten carbide ball in air across an En 352 surface part of which had been implanted with about 3×10^{16} Sn ions/cm² at an energy of 380 KeV. En 352 is a case-

hardening steel containing about 4 atomic percent of other metals. The coefficient of friction decreased sharply from 0.24 to 0.09 at the boundary between the unimplanted region and the implanted region, and then increased sharply back to the value 0.24 at the other boundary.

When the steel surface was implanted with In, Mo, S or Kr ions, little or no change in the coefficient of friction was observed. When a steel surface was implanted with both Mo and S ions in the ratio 1:2, a decrease of about 30% was observed for the coefficient of friction. Surfaces implanted with Ag or Pb ions showed small increases in the coefficient of friction.

Of course there are several potential factors that could conceivably have affected the result other than the direct presence of the implanted ions. For example, when ion bombardment occurs, there is a tendency for pump-oil molecules in the vacuum to be cracked, resulting in the deposition of carbon in the bombarded region. But precautions were taken to preclude this condition. Another possibility is the effect of surface radiation damage on the coefficient of friction; but the absence of any change in the coefficient of friction when the sample was implanted with Krypton implies surface radiation damage was not responsible for the change in the coefficient of friction.

EXPERIMENTS ON WEAR

In the April issue of Tribology Magazine, Hartley² of the University of Sussex published the results of a series of pin-and-disk wear tests. He used an Avery-Denison T62 Tribotester, and compared the Archard wear parameter

$$K = \frac{3H\Delta V}{LX} \quad (1)$$

for various implanted surfaces with the unimplanted wear parameter. In Eq. (1) H is the surface hardness, ΔV is the volume removed by wear, L is the load on the pin, and X is the distance traveled. The implanted surface was the top of the disk, which was rotated under a loaded stainless steel pin with a conical end in contact with the disk. For the larger loads, (about 40 newtons) the wear parameter was roughly an order of magnitude lower for the surface implanted with nitrogen ($\sim 10^{18}$ ions/cm² at 35 KeV) than for the unimplanted surface.

A separate implantation with Mo ions (about $2.8 \times 10^{16}/\text{cm}^2$ at 400 KeV) gave similar results. In this case the concentration of implanted atoms was substantially less than in the former case because fewer ions/cm² were implanted and also because the ions are distributed over a greater range in depth (because of the greater implantation energy).

Another steel disk was implanted with about 1.8×10^{18} N⁺ ions/cm² at an energy of 30 KeV, and the wear parameter was measured as a function of sliding distance from about 10^2 m to about 4×10^4 m, with a constant load of about 15 N. The differ-

ence between surface wear for implanted and unimplanted conditions was negligible for a distance of 10^3 m and gradually increased as a function of distance to about an order of magnitude at the maximum distance observed.

RELATIVE DISADVANTAGES AND ADVANTAGES OF ION IMPLANTATION

The relative disadvantages of the use of ion implantation to change the properties of a surface as compared with more conventional methods are listed in Table I. These characteristics are not always disadvantages. For example, in the application of ion implantation to the problem of wear, radiation damage (which produces large compressive stresses at the surface) might be an advantage, and the configuration limiting aspect of ion implantation is not always undesirable; it enables one to make an implantation at a precisely defined location. Even sputtering of the surface is not always bad; it tends to clean the surface. The limitation of ion implantation to thin surface layers is sometimes desirable because a good surface characteristic (e.g., noncorrosion) is sometimes correlated with an undesirable bulk characteristic (low strength). For some purposes (e.g., preparation of non-equilibrium concentrations of different materials) ion implantation can be much less expensive than conventional metallurgical techniques.

Table II lists the advantages of ion implantation in modifying surface characteristics.

TABLE I.

DISADVANTAGES AND LIMITATIONS

1. Radiation Damage to Host Material
2. Configuration Limited (Surface must be exposed to vacuum during irradiation)
3. Sputtering of Target Surface
4. Limited to Thin Surface Layers
5. Relatively Expensive

TABLE II.

ADVANTAGES OF ION IMPLANTATION

1. No Change in Dimensions
2. No Interfaces
3. No Coatings to Peel Off
4. All Elements and Combinations of Elements
Are Possible
5. Depth Concentration Distribution Controllable
6. No Sacrifice of Bulk Properties
7. No Problems with Thermodynamic Diffusivity
8. Composition May Be Changed Without Affecting
Grain Sizes
9. Precise Location of Implantation

Basically, ion implantation gives us a brute force method of introducing atoms of arbitrary species to an arbitrary host material in arbitrary relative concentrations; i.e., this nonequilibrium process is not limited by the ordinary laws of thermodynamics and physical chemistry, such as solubilities and diffusivities.

CONCLUSION

The friction and wear experiments discussed herein are not definitive. There are many parameters involved, and more comprehensive measurements must be made before definite conclusions can be drawn as to the usefulness of implantation in improving the wearability of surfaces. There appears to be a sufficiently large number of investigations now underway that such definite conclusions can be drawn within a few years.

References

1. N. E. W. Hartley, G. Dearnaley and J. F. Turner, Proc. Int. Conf. Ion Implantation Semicond. Other Materials 3rd, Yorktown Heights, New York, 1972, Plenum Press, New York.
2. N. E. W. Hartley, Tribology International, April 1975, p. 65.

COMMENTS ON THE BOWDEN AND TABOR

ASPERITY THEORY OF WEAR

Vernon Westcott

The most commonly discussed theory of wear is that presented by Bowden and Tabor in the 1930's. The process of wear is described as the interaction of asperities on opposing surfaces in which part of one asperity is sheared off and sticks to the opposing surface.¹

A fraction of such particles separate from the opposing surface and become free wear particles which will enter the lubricant if present.

The wide acceptance of this theory seems to have been the result of its explaining Aminton's law, i.e., the proportionality between the friction force and the perpendicular force driving the surfaces together.

However, when the theory is applied to microscopic details of wear it does not adequately explain a number of known factors.

Our objections (Scott, Wright and Westcott) are the following:

The theory treats the generation of a wear particle as a single event. The wear particle is visualized as the result of shearing off of a part of an asperity. The effect

of the previous wear history is considered only in terms of the wear having modified the surface, i.e., the population and size of the asperities.

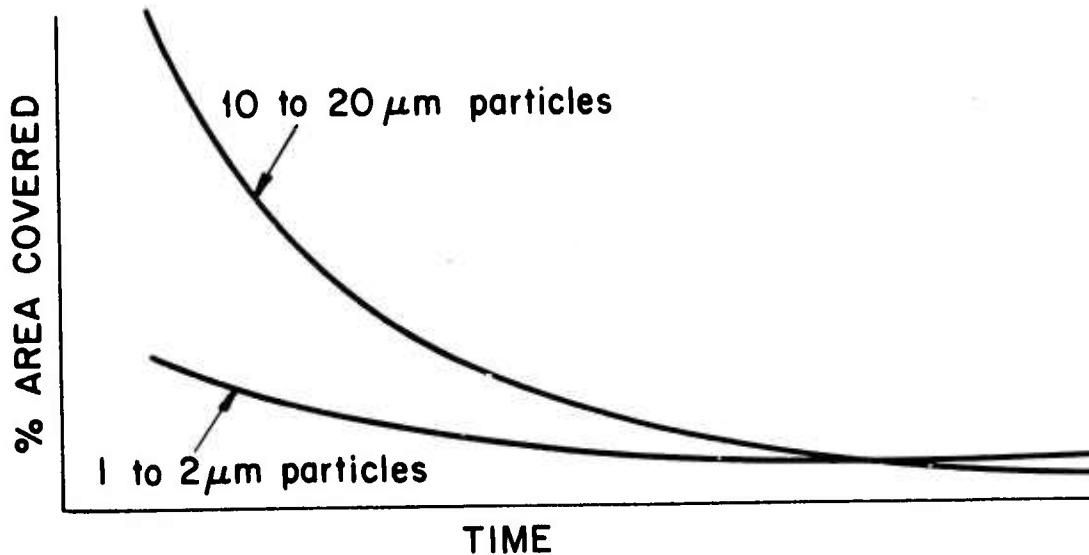
On the other hand, the evidence from particle analysis and observation of worn surfaces shows that very low sliding wear rates are accompanied by the presence of a shear mixed layer.

The asperity theory does not account for the shear mixed layer or explain the way in which the surfaces "break in".

When two relatively smooth (commercially machined) surfaces slide in the presence of a lubricant large wear particles are generated initially and the maximum size of these particles decreases with further running. The initial particles appear jagged, may for example, be 150 μm long and are the result of breaking off of surface irregularities such as the tops of grinding marks.

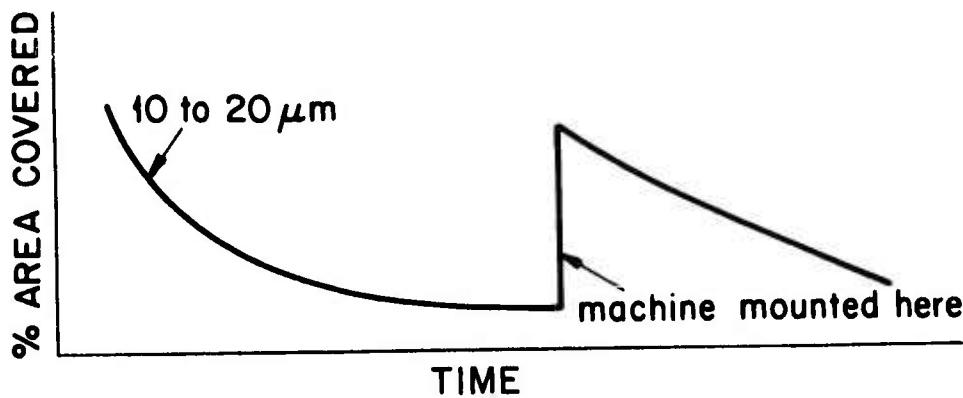
In some cases of sliding steel surfaces such as 52100 or 1018 steel, minute steel spheres (which are characteristic of micro cracks in the surface of cyclically loaded surfaces) are found in small quantities. As the surfaces continue to rub the generation of the spheres stops.

The generation rate of wear particles and their size decreases with time along an exponential curve.



The ratio of percent area covered by the large and small particles in the above graph approaches unity and the generation rate may be down by a factor of 100 or more.

Any disturbance of the operating conditions causes the large particles to reappear and if the disturbance is eliminated the surfaces will again "wear in".



The asperity theory of Bowden and Tabor does not treat the wear regimes in which the principal type of wear particle

is an oxide.

When steel surfaces slide over each other, it is possible by the proper selection of operating parameters, for the sliding surfaces to generate wear particles which are predominantly of a single type of oxide.

For example, in regime 4² most of the wear particle volume is composed of alpha hematite, Fe_2O_3 . It does not appear that the mechanism of generation of the resulting large polycrystalline hematite particles can be described as the breaking off of an asperity.

References

1. F. P. Bowden and D. Tabor, The Friction and Lubrication of Solids, Clarendon Press, Oxford, Part I 1954, Part II 1964.
2. A. A. Reda, R. Bowen and V. C. Westcott, "Characteristics of Particles Generated at the Interface Between Sliding Steel Surfaces," Wear 34, 261-273 (1975).

COMMENTS ON WEAR PARTICLE SIZES

Vernon Westcott

The following comments apply to the wearing of steel such as 52100 or 1018. The maximum particle size for a low (benign wear mode) wear rate is in the order of $15\mu\text{m}$. Wear regimes 1 and 2 produce such particles.

If the shear mixed layer is destroyed a much higher wear rate results. The increase may be 10^2 to 10^4 . The maximum particle size is variable depending on the material, load, etc., but particles of 100 to $150\mu\text{m}$ are not uncommon. When wearing in this higher wear mode, Regime 3, most of the metal volume is contained in the larger particles >80%.

FUNDAMENTALS OF WEAR

A. H. Shabaik

The behavior of materials in wear situations depends upon the type of contact, the level of stress and temperature, and the nature of the environment. The basic types of wear found in the majority of engineering situations are:

1. Adhesive wear
2. Abrasive wear
3. Fretting wear
4. Erosive wear
5. Corrosive wear
6. Combination of the above (e.g., adhesion-abrasion, fretting-corrosion, erosion-corrosion)

A general wear mechanism may be difficult to propose, however, the essential elements of the wear process may be classified as:

1. Surface interaction
2. Plastic deformation
3. Crack initiation, fracture and separation

Combinations of these in varying degrees can account for the different types of wear.

Factors that affect the wear process are metallurgical (mechanical and physical properties of the surface: hardness, toughness, microstructure, and composition), operational (load, speeds, types of motions, and environment), geometrical (microtopography, asperity distribution, size and shape). A

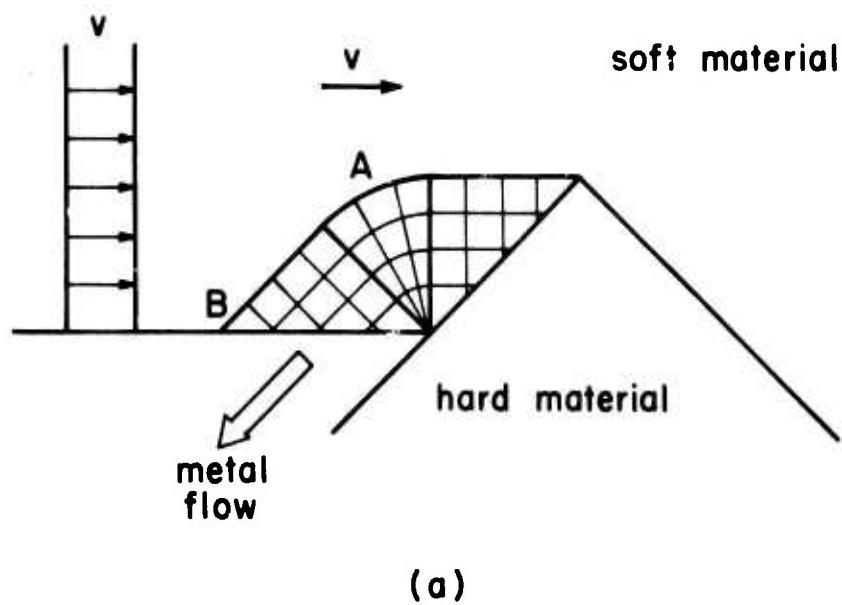
tribological approach to examine the effect of these factors and its relative importance on wear behavior of materials is essential for wear control.

The mechanism of metal deformation (on a microscale) and debris formation during sliding of both flat and wedged hard surfaces is shown in Fig. 1. In the first case (Fig. 1a) the metal will pile up in front of the slider and finally shears off at the elastic-plastic boundaries. Once the wear particle is separated the new boundary AB will form the new contact area and the process of deformation and separation will continue. The wear debris generated in this case is mostly of a small size particle.

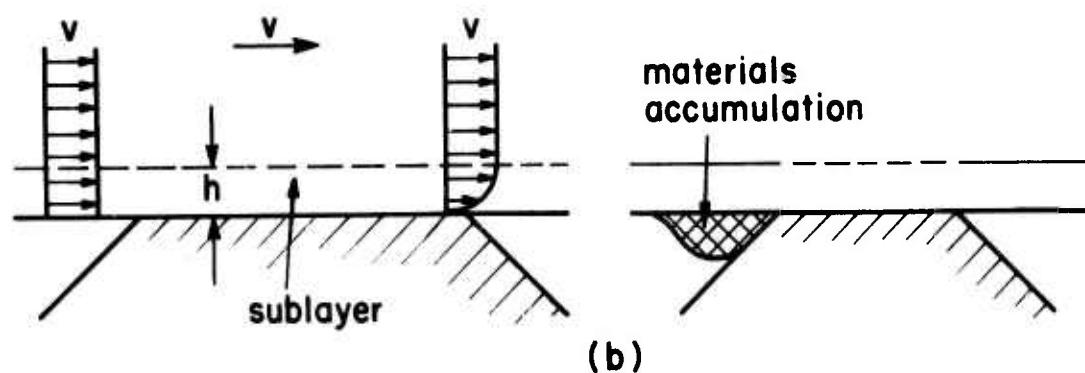
In the second case (Fig. 1b), materials accumulation as a result of the nonuniformity of the velocity distribution in the sublayer of the soft material will take place and subsequent shearing off of the boundary layer and/or separation of the material at the end of the slider will occur. The wear debris will mostly be of plate like particles with a mixture of small particles.

Both previous models can operate under the same sliding situation. The degree of each will depend on the surface topography of the hard slider.

Experimental examination of wear debris during sliding (Fig. 2) shows a mixture of equiaxed, microchip, and plate like particles.



(a)



(b)

Figure 1. Material flow rate entering contact area $Q_I = v \cdot h$,
material flow rate leaving contact area $Q_O = \int_0^h v dy$,
 $Q_I > Q_O$.

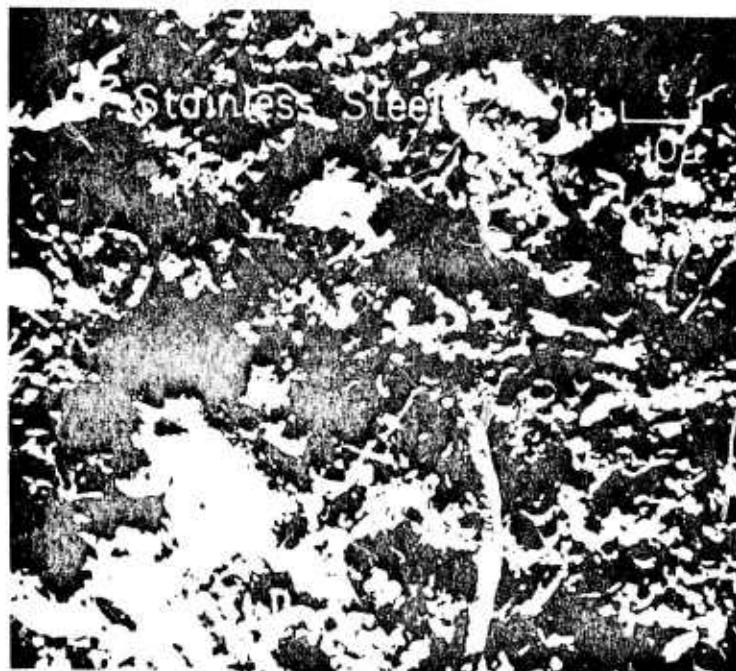
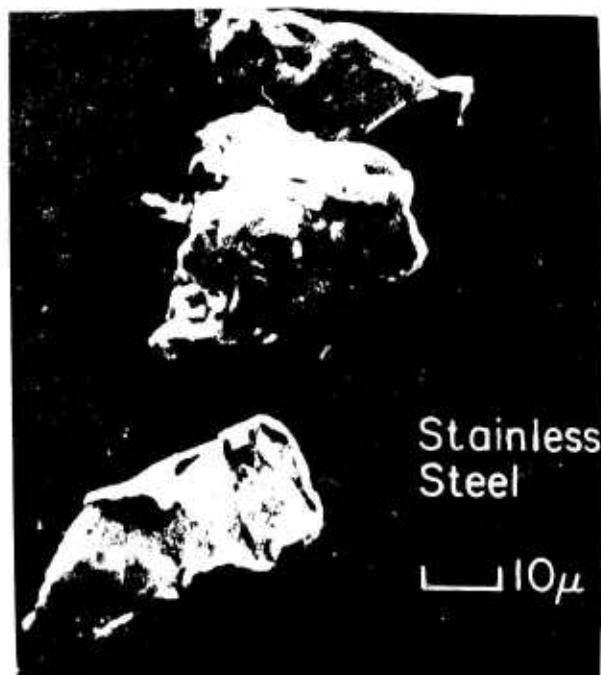


Figure 2. Experimental examination of wear debris during sliding.

In fretting wear, spherical particles have been observed by a number of investigators. Figure 3 shows such an example. The high degree of surface deformation and the preferential direction along which crack propagation takes place can be seen. Lifting off of the heavily deformed surface can result from loose particles, hard spots extending from the contacting surface, and/or fatigue of the surface layer.

Examination of the effect of the metallurgical, operational, and geometrical parameters on the wear behavior and wear mechanism of solids is very important and does not need to be emphasized. The information obtained will provide an important step in both wear control and designing systems with a reasonable service life.

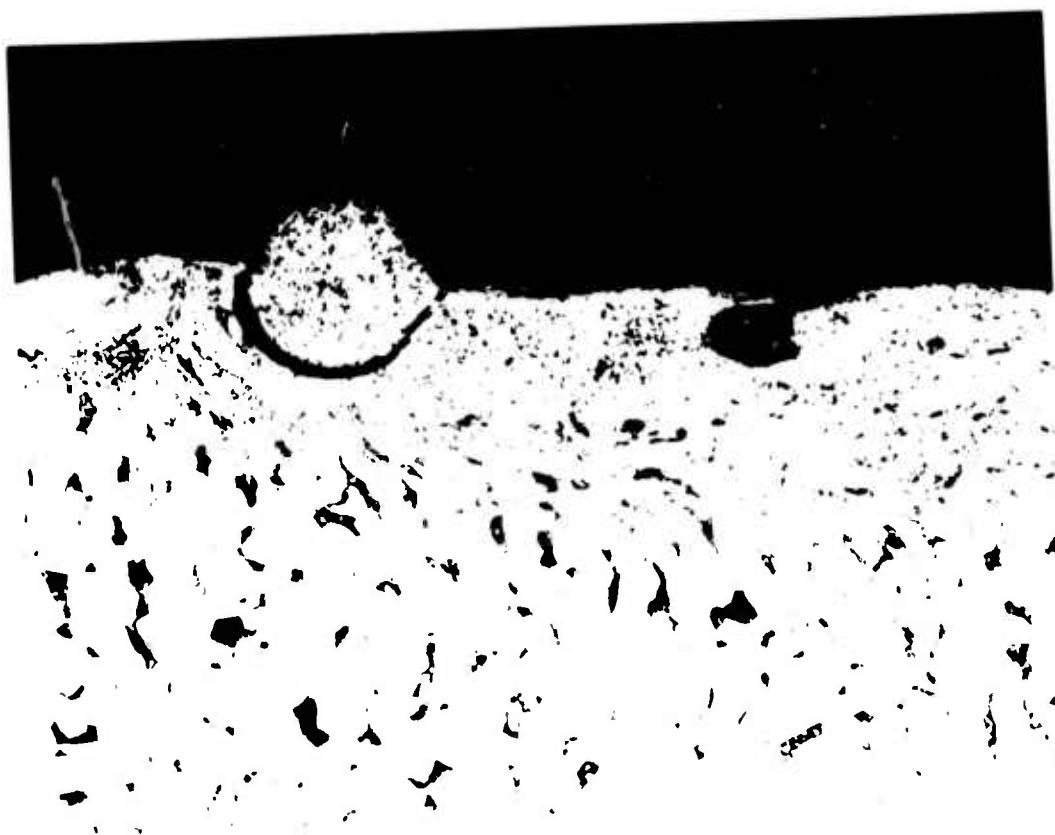


Figure 3. Metallographic section through fretted surface:
72,000 cycles of fretting: 500 C. ($\times 150$)

ON THE DELAMINATION THEORY OF WEAR

N. P. Suh

The delamination theory of wear states that wear particles are created by the following process:

1. When two sliding surfaces come in contact, normal and tangential loads are transmitted through contact points. Small asperities of the softer surface are easily deformed and fractured by the repeated loading action, forming small wear particles of various shapes and eventually creating a smooth surface. At the same time bulk deformation of the surface layer takes place due to the surface traction. (Note that if the removal of asperities is the only wear mechanism, wear should cease once a surface becomes smooth.)

2. Plastic deformation is due to both the motion and the generation of dislocations. However, dislocations very near a surface are subjected to a sufficient image force to be pulled out of the surface as the coherent oxide layer is broken by the action of moving hard asperities. If a strong coherent oxide layer with a modulus higher than the substrate is present, the dislocations will be pushed away from the interface into the substrate. The dislocations below a critical depth, where the image force is not strong enough, are stable, and as they accumulate they work-harden the sub-surface.

(The dislocation behavior in this region needs further clarification.) Consequently, the deformed material exhibits a flow stress gradient near the surface: as the distance from the surface increases the hardness initially increases, reaches a maximum value, and then decreases.

3. A consequence of the flow stress gradient is that the hard asperities more easily "penetrate" the surface of the softer metal, displacing and spreading the softer metal at the surface. Therefore, the surface traction is caused by both the plowing of the softer surface by hard asperities as well as by the adhesive force. Furthermore, once the major asperities are removed, the contact is not just an asperity-to-asperity contact but rather the surface traction acts all along the softer surface. The plowed material may also be removed in a manner similar to the removal of asperities.

4. As the sub-surface deformation continues, cracks are nucleated away from the surface. Crack nucleation very near the surface is not favored for two reasons: a triaxial state of compressive stress existing just below the contact points, and the existence of the "soft" metal near the surface, which does not work-harden appreciably and thus prevents stress concentrations by maintaining a low flow stress. Crack nucleation is preferentially initiated at particle-matrix interfaces of hard inclusions.

5. Upon further deformation, these cracks extend and propagate, joining neighboring cracks. This propagation is

controlled by the tensile state of loading existing near the surface over which the asperity has just passed by. These cracks tend to run parallel to the surface since the loading condition is the same all along the surface.

6. When these cracks finally reach the surface, thin, long, wear sheets delaminate. These delaminated wear particles have dominating influence over the life of a part rather than the wear particles generated by deformation of surface asperities.

We believe that the experimental and theoretical work done at MIT and elsewhere (NASA, Nottingham, Japan, the Netherlands, etc.) supports the validity of the delamination theory of wear. We are in the process of quantitatively analyzing the delamination process. This is expected to be a three-year project.

One of the practical results of the delamination theory is the development of a method of minimizing sliding and fretting wear by plating the surface of a hard substrate with a metal having lower flow stress, such as nickel, silver, gold, or cadmium. When the thickness of the surface layer is thinner than a critical value, the layer can deform continuously, since it does not work-harden and remains soft. Therefore, the tangential load transmitted to the substrate is substantially minimized. Since the wear of metals between two sliding surfaces at low speeds is greatly influenced by the tangential component of the surface traction, the wear rate can accordingly be reduced.

The thickness of the "soft" layer associated with dislocation instability (i.e., non work-hardening layer) postulated by the delamination theory of wear is not the same as the thickness of delaminated wear sheets. It is a fraction of a micron in the case of most f.c.c. metals and at least an order of magnitude smaller in the case of b.c.c. metals. The thickness of the wear sheet is determined by the depth of the crack which is controlled by the state of loading for a given metal.

ON THE EXPERIMENTAL RESULTS OBTAINED
AT THE TRANS-SONICS, INC.

The wear particles in Regime 1 (of Mr. Westcott's wear particle classification) is also created by the delamination process. The reason they are thinner than the wear sheets shown in the MIT reports are that

1. Mr. Westcott used 52100 steel which was much harder (i.e., $H_B = 700 \text{ kg/mm}^2$) than the specimens used at MIT (the hardness of the MIT specimens ranged from $35-180 \text{ kg/mm}^2$).
2. Furthermore, the microstructure of 52100 steel is quite different from those shown in the MIT report. The 52100 steel used has a large number of second phase particles and therefore would not show the same fracture surface. The details of these surface wear particles cannot be resolved by the optical microscopy used at Trans-Sonics, Inc.

There is no concrete evidence that the Beilby layer (i.e., the shear mixed layer of Mr. Westcott) is any different

from any other crystalline metal. The claim that the Beilby layer is amorphous cannot be substantiated by scientific evidence.

After I came back to MIT from the workshop, I called Mr. Ahmed Reda, the engineer at Trans-Sonics, Inc., who obtained the results presented by Mr. Westcott. The experimental conditions supplied by Reda over the telephone are given in tabular form in Table I. Please note the following:

1. In all regimes a fixed normal load was applied.

Therefore, the thickness given by Mr. Westcott was for a given set of loads on 52100 steel against 52100 steel. 52100 steel has a hardness of about 700 kg/mm^2 . The observed thin wear sheets are to be expected when hard materials are loaded moderately.

2. Mr. Reda said that as soon as the normal load was increased, the thickness increased as shown for Regime 3. This demonstrates that wear sheet thickness is a function of normal load as predicted by the delamination theory of wear.

3. Regime 6 is so severe that there must have been a large temperature rise at the interface. This would have transformed martensite in the subsurface of 52100 bearing materials. The delamination theory of wear has not been explored in materials experiencing phase transitions.

4. In none of their experiments was tangential load measured in spite of its importance in determining failure mechanism.

TABLE I - EXPERIMENTAL CONDITIONS (supplied by Mr. Ahmed Reda*
of Trans-Sonics, Inc.)

	Material Tested	Normal Load	Sliding Velocity	Interfacial Condition	Thickness of Wear Sheet
Regime 1	52100 on 52100 steel	11 Kg	0.76 m/sec	lubricated	1 micron
Regime 2	52100 on 52100	35 Kg	0.76 m/sec	lubricated	1 micron
	52100 on 1018	18 Kg	0.76 m/sec	lubricated	1 micron
Regime 3	52100 on 52100	180 Kg	0.47 m/sec	lubricated	5 micron
Regime 4	52100 on 52100	1.1 Kg	0.47 m/sec	unlubricated	
Regime 5	52100 on 52100	33 Kg	0.47 m/sec	unlubricated	
Regime 6	52100 on 52100	1.1 Kg	5m/sec	unlubricated	10 to 20 micron
	52100 on 52100	11 Kg	2.2 m/sec	unlubricated	

*Mr. Reda is the engineer who is in charge of the wear project
at Mr. Westcott's company.

5. Mr. Reda states that with 1018 steel it was difficult to go into Regime 6 as would be expected if Regime 6 is associated with phase transition.

ON MIT'S FUTURE WORK

The objectives of the MIT project are:

1. The development of a quantitative method of determining wear through further experimental and theoretical analysis,
2. Understanding of the behavior of materials near the surface,
3. Innovation of methods of preventing wear.

To reach these objectives the following needs to be done:

1. The analysis of:
 - a. stress field
 - b. deformation field
 - c. crack nucleation
 - d. crack propagation
2. Low cycle fatigue tests to determine:
 - a. cyclic behavior of materials
 - b. $\frac{dc}{dn}$ (crack growth rate)
3. Further investigation of the role of thin soft metal layers in preventing fretting wear and fretting fatigue.
4. The effects of metallurgical variables on wear.
5. Systematic investigation of exfoliation of worn surfaces by chemical reaction with lubricants to establish the criteria for acceptable chemical components.

6. Research on the dislocation structure near the surface to further clarify surface structures and properties and to enable us to design a better wear resistant surface than hitherto possible.

7. Extension of the delamination theory of wear to situations where chemically and thermally activated processed also play a role.

The availability of research funds will determine the scope of the MIT project. During the period of 1975-1976 the items (1) through (5) of the work to be done will be investigated.

There are many other questions which need to be answered. It is my hope that many other scientists and engineers will initiate research work in order to develop a concrete scientific foundation for this field which has been dominated by empiricism.

FATIGUE ANALYSIS OF WEAR

B. Budiansky and F. A. McClintock

ABSTRACT

There appears to be a moderate amount of credence in the validity of Archard's adhesive wear formula

$$\left(\frac{\text{Wear Volume}}{\text{Sliding Distance}} \right) = \frac{k}{3} \left(\frac{\text{Normal Force}}{\text{Hardness}} \right) \quad (1)$$

in which the factor k , to be determined experimentally, is described as a rip-off probability associated with an adhesive encounter of opposing asperities. Suh's delamination theory of wear rejects the notion that wear occurs by the knocking-off of asperities, but rather results from successive delamination of thin sheets of material. In the present, admittedly crude, analysis we reconcile Archard's formula with Suh's mechanism, and discover that k can then be interpreted as a number inversely proportional to the number of cycles of sliding contact load needed to produce a fatigue fracture just below the surface of the material, in the vicinity of an asperity.

FATIGUE ANALYSIS OF WEAR

B. Budiansky and F. A. McClintock

Archard's formula for wear rate¹ is

$$\frac{V}{S} = \frac{kL}{3H} \quad (1)$$

where V = volume of wear material

S = total sliding distance

L = normal load

H = hardness (indentation pressure)

and k is a non-dimensional number to be determined by experiment. If two dissimilar materials are involved in the sliding contact, the hardness in (1) is supposed to be that of the harder material. In Archard's derivation, k is described as the probability of detachment of an asperity during an adhesive encounter of opposing asperities. Substitution of test results into (1) provides values of k as high as 10^{-2} for very severe wear, while "benign" wear is associated with k as low as 10^{-8} . The constant k is supposed to depend only on the identities of the materials in contact and the state of lubrication, but we are unaware of systematic experiments that show the extent to which k is independent of the normal loading, since the experiments of Rabincowicz and Tabor (1951) and Burwell and Strang (1952), who varied loads by a factor of about 10. They found

k to remain constant as long as the nominal normal contact stress was less than about one-third the hardness.

A brief review of Archard's derivation of (1) will facilitate further discussion. Under the normal load L , the actual area of contact A_c is estimated crudely as*

$$A_c = L/H \quad (2)$$

If this area is imagined to consist of n disjoint regions of roughly circular shape of average diameter d , then, approximately,

$$A_c = \pi d^2 n/4 \quad (3)$$

Attention is now focussed on the fate of a given asperity. During an adhesive encounter with an asperity of the opposite face, its contact area will vary; but the average diameter of this area during the encounter is approximately equal to d ; and the sliding distance during which the asperity is contacted is also roughly d . Consequently, about n/c new contacts are made (and an equal number unmade) per unit sliding distance. If a fraction k of these adhesive contacts results in fracture and breaking off of v units of material volume per fracture, the volume lost per unit sliding distance is

*Subsequent work by A. P. Green (1954) has shown that for simultaneous sliding and relatively flat asperities, the effective hardness drops to $H/2$. With finite angles the effective hardness would be even less, perhaps corresponding to the $H/3$ threshold mentioned above.

$$\frac{V}{S} = k(n/d)v \quad (4)$$

Now let v be the volume of a hemisphere of diameter d :

$$v = \pi d^3/12 \quad (5)$$

Then

$$\frac{V}{S} = \frac{\pi}{12} kd^2n \quad (6)$$

But, by (2) and (3)

$$d^2 = \frac{4L}{\pi Hn} \quad (7)$$

and Eq. (1) follows.

We come now to the essential point of this note, which is to indicate that if wear is actually a fatigue process, the factor k may be related to a characteristic fatigue lifetime of N cycles appropriate to repeated application of the loads experienced by the surface asperities. Suppose that N repetitions of a typical cycle of normal and tangential stress applied to an asperity would lead to a fracture that produces a volume loss v . Then a large number $M \gg N$ of such loading incidents, applied to asperities, would give a volume loss approximately equal to Mv/N . But we estimated earlier that there are n/d such encounters per unit sliding distance, hence $M = nS/d$, and so

$$V/S = \left(\frac{nv}{Nd} \right) \quad (8)$$

Accordingly, the Archard constant k equals $(1/N)$.

If we suppose that material removal is in the form of hemispherical chunks, as before, the formula (1) is obtained, with $k = 1/N$. But there is no need to retain this assumption. We can, for example, assume, in accordance with Suh's "delamination" theory, that repeated loading of asperities leads to material loss in the form of thin flakes from the vicinity of the asperity. Then, even if material parameters set the depth of fracture, a constant geometry of cracking sets the volume removed to be proportional to the cube of the asperity dimension, say the diameter d :

$$v = \alpha d^3 \quad (9)$$

where α is a characteristic dimensionless factor. Introducing (9) into (8),

$$\frac{v}{s} = \frac{\alpha d^3}{N} \quad (10)$$

and then (7) gives

$$\frac{v}{s} = \left(\frac{4\alpha}{\pi N} \right) \left(\frac{L}{H} \right) \quad (11)$$

We can not, of course, pretend that (11) is precise or even very useful as it stands. It does seem, however, to suggest that an appropriate avenue for future attempts to develop a rational, quantitative theory of wear may well focus on the roles played by the details of repeated stresses in the vicinity of loaded asperities, and their relations to fatigue crack growth rates and fatigue failure. The large shear

deformations observed by Suh and others probably arise from cycle dependent flow under a mean shear stress.

Furthermore we do not claim that the idea of wear as a micro-fatigue process is novel. Professors Rabinowicz, Suh, and others have kindly provided references including Bayer et al. (1962), Kraghelsky (1965), Bayer and Schumacher (1968) (who also showed micrographs of extensive deformation and cracking on the surface), and Kimura (1974).

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

References

Archard, J. F. 1953 "Contact and Rubbing of Flat Surfaces," *J. Appl. Phys.* 24, 981-988.

Archard, J. F. 1957 "The Wear of Metals under Unlubricated Conditions," *Proc. Roy. Soc. A* 236, 397-410.

Bayer, R. G. 1962 "Engineering model for wear," *Wear* 5, 378-391.

Clinton, W. C.

Nelson, C. W.

Schumacher, R. A.

Bayer, R. G. 1968 "On the Significance of Surface Fatigue in Sliding Wear," *Wear* 12, 173-183.

Schumacher, R. A.

Burwell, J. T. 1957 "Survey of Possible Wear Mechanisms," *Wear* 1, 119-141.

Burwell, J. T. 1952 "On the Empirical Law of Adhesive Wear," *J. Appl. Phys.* 23 18-28. See also, Burwell, 1957, and Archard and Hirst, 1957.

Strang, C. D.

Green, A. P. 1954 "The Plastic Yielding of Metal Junctions due to Combined Shear and Pressure," *J. Mech. Phys. Solids* 2, 197-211.

Kimura, Y. 1974 "An Interpretation of Wear as a Fatigue Process," *Apparently Some Japanese Proceedings*, 89-95.

Kraghelsky, I. V. 1965 "Calculation of Wear Rate," *J. Basic Eng. Trans. ASME*, V.87D, 785-790.

Rabinowicz, E. 1951 "Metallic Transfer between Sliding Metals: An Autoradiographic Study," *Proc. Roy. Soc. (London)* A208, 455-475.

Tabor, D.

ELEMENTARY RESULTS OF DIMENSIONAL ANALYSIS FOR RATES
OF WEAR IN STEADY STATE SLIDING

D. C. Drucker

ABSTRACT

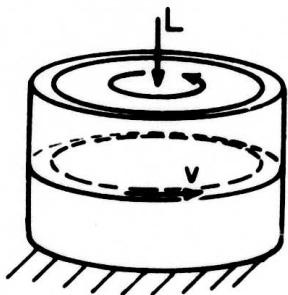
The combination of a dimensional analysis and the recognition of surface roughness (asperity) dimensions as dependent rather than independent variables in a steady state of wear leads to some interesting results. Functional forms for the friction coefficient, wear rate, and the geometry of the wear surface are seen to provide surprisingly strong restrictions in some instances and surprisingly little in others. The role of load, rubbing speed, elastic and plastic properties of rubbing surfaces, surface energy, critical stress intensities, microstructural dimensions, heat conductivity and strain rate sensitivity are touched upon briefly.

ELEMENTARY RESULTS OF DIMENSIONAL ANALYSIS FOR RATES
OF WEAR IN STEADY STATE SLIDING

D. C. Drucker

Introduction

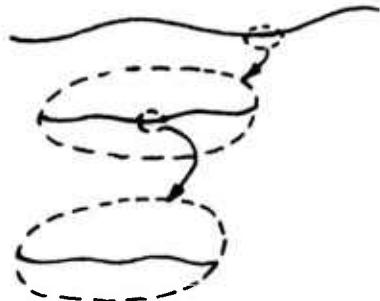
A dimensional analysis cannot give an explicit analytical solution to a problem of mechanics such as the rate of wear in sliding. However, it can and does serve an extremely useful purpose in advance of or conjunction with a theoretical analysis or an experimental program. The listing of all the independent variables thought to enter into a mathematical solution or an experimental result often leads to surprisingly strong limitations on possible answers and great improvement in intuition.



To be specific, consider the equivalent of an infinitely long slider in the form of a ring test. Suppose, as is generally reported, that ex-

periments show a steady wear rate to be achieved after an initial (run-in) period. Equivalently, imagine that a steady wear rate is assumed in the development of a theory of wear. In the steady state, the rubbing surfaces established by the process of wear are statistically similar over time. The

irregularities or asperities on all scales remain the same, as do the contact and subsurface stresses, again in statistical average.



Macroscopic Variables Only

Consider first the case for which an assumption is made explicitly or implicitly that the only independent variables are the macroscopic ones listed below:

v	the rubbing speed
L/A	the average or nominal normal stress on the surface of sliding
σ^Y, n	the yield stress and strain hardening exponent symbolizing the inelastic stress-strain behavior
E, ν	the moduli of elasticity

Should the upper and lower rings be of different material, the inelastic and elastic variables of both would be included and appear in ratio form as $\sigma_{\text{upper}}^Y/\sigma_{\text{lower}}^Y$, $E_{\text{upper}}/E_{\text{lower}}$, etc. Similarly, σ^Y should be thought of as also representing symbolically each significant stress value σ^s . Values for cyclic loading and complex loading as well as monotonic may be included. With this understanding, all S-N fatigue points or curves can be included along with cyclic hardening or softening curves, etc.

Each such stress value (conceptually there may be infinitely many sets of infinitely many values each) enters in ratio σ^s/σ^Y . In what follows these ratios along with $\sigma_{upper}^Y/\sigma_{lower}^Y$, etc. will be implicit but not written down explicitly.

Dependent variables of interest are:

f	the coefficient of friction
$dD_w/dt = (D_w/t)$	the rate of wear (thickness or depth D_w of material removed per unit of time)
δ_s	any one of the typical asperity dimensions of the surface or equally well the thickness of a delaminated wear particle δ

Dimensional analysis requires that a dimensionless form of the dependent variable be a function of the independent dimensionless combinations of the independent variables. For example, any analytical solution to the problem of wear as posed must be of the form f and $(D_w/t)/v$ are functions of σ^Y/E , $(L/A)/\sigma^Y$, n , v . If any experimental (real) result contradicts this functional form, the conclusion must be that the listing of important independent variables is not complete.

Note that with independent variables of the class v , (L/A) , σ^Y , n , E , v there is no natural length or time resulting from any combination of them. Distance slid or elapsed time can play no real role in the steady state beyond the statement that D_w is linear in either. Three important consequences follow immediately from the absence of a length and time scale.

The first is that the coefficient of friction f is independent of the rubbing speed v , and the second is that the wear rate is proportional to the rubbing speed as already written down.

The third is that each asperity dimension δ_s is zero or irrelevant because no meaningful dimensionless combination of δ_s and the independent variables exist. Similarly the thickness δ of a delamination wear particle is undetermined.

Length Scale

If δ_s or δ is to be computable by the simplest or most elaborate theory, the list of independent variables must be expanded. Metallurgical lengths, l_m , such as grain size or major inclusion size or spacing should be considered. However, within homogenous continuum mechanics, adding either

γ^A an interfacial surface energy per unit area

or

K_c a critical value of fracture parameter with units of stress times the square root of length

introduces a choice of natural lengths in addition to the l_m .

One such choice for γ^A is γ^A/σ^Y ; one such choice for K_c is $K_c^2/(\sigma^Y)^2$. The choice of dimensionless variable is not unique. Obviously, E or $\sqrt{E\sigma^Y}$, or ... could be substituted for σ^Y . With γ^A alone (no metallurgical length l_m) f , $(D_w/t)/v$ and $\delta_s/(\gamma^A/\sigma^Y)$ are functions of $(L/A)/\sigma^Y$, n , σ^Y/E , v ; γ^A cannot enter into the coefficient of friction or the rate of wear.

Similarly if K_c alone is added to the independent variables, the rate of wear cannot depend upon K_c .

If both γ^A and K_c are added, only one and not two new independent dimensionless variables can enter

f , $(D_w/t)/v$ and $\delta_s/(\gamma^A/\sigma^Y)$ or $\delta_s/[K_c^2/(\sigma^Y)^2]$ or other dependent dimensionless variables such as $\delta_s/[K_c/\sigma^Y]\sqrt{(\gamma^A/\sigma^Y)}$ must now be functions of $(L/A)/\sigma^Y$, n , σ^Y/E , v , $\gamma^A\sigma^Y/K_c^2$.

When a metallurgical length ℓ_m is introduced, then the more reasonable result is that f , $(D_w/t)/v$ and δ_s/ℓ_m or $\delta_s/(\gamma^A/\sigma^Y)$ may be functions of $\gamma^A/\sigma^Y\ell_m$ in addition to the previous dimensionless independent variables.

Time Scale

Natural lengths have been introduced but no natural time. Wear rate must remain linear in v , and f must remain independent of v .

Variation with v requires further addition to the independent variables. Suppose the materials still to be time independent (no creep or relaxation) but heated sufficiently in the steady state to alter their yield stress or other material properties. The steady state temperature distribution depends upon the heat conductivity k which must be added to the independent variables along with such quantities as a coefficient b of the rate of variation of σ^Y with temperature; $\sigma^Y = \sigma_0^Y = b\Delta\theta$, where σ_0^Y is the yield stress at some reference temperature θ . In the absence of the influence of temperature on material properties, k would be without influence on the

coefficient of friction or on the rate of wear. With a significant influence on the materials, an additional independent dimensionless variable enters the functional form for f ,

$(D_w/t)/v$ or $\delta_s/(\gamma^A/\sigma_0^Y)$. One permissible choice is $(vb/k)(\gamma^A/\sigma_0^Y)$ which does in fact lead to variation of the coefficient of friction and to nonlinearity of the wear rate with v .

If, instead, the effects of temperature change are or are assumed to be negligible but the material is strain-rate sensitive, say, symbolized by the one dimensional form, $\dot{\epsilon}/\dot{\epsilon}_r = (\sigma/\sigma_r)^m$ then f , $(D_w/t)/v$, and $\delta_s/(\gamma^A/\sigma_0^Y)$ are functions of $(L/A)/\sigma_0^Y$, $\gamma^A/\sigma_0^Y \dot{\epsilon}_m$, $\gamma^A \dot{\epsilon}_r/\sigma_0^Y v$, $\gamma^A \sigma_0^Y / K_c^2$, σ_0^Y/E , v , n , σ_r/σ_0^Y , m

material properties in dimensionless form

and variation of f or nonlinearity of wear rate with v can be appreciable.

Concluding Remarks

In summary, the dimensionless groupings say a great deal about the form of result that an analysis can produce from a given set of mechanics variables or the variables that are needed to explain a given set of experimental results. Dimensional analysis, however, falls far short of providing the result. For example, the reported linearity of rate of wear with normal load L is not a consequence of even the very first and simplest choice of independent macroscopic variables. Also, dimensional analysis alone provides no guidance on the

most significant of or the best practical choice among the infinitude of permissible dimensionless variables. It does not even provide a check on the reasonableness of the assumption of a steady state as distinguished from a graduate or abrupt transition between an initial or run-in period and subsequent catastrophic wear.

Very Recent References of Interest

Papers presented at the Third International Tribology Conference, "Tribology for the Eighties," September 22-25, 1975 and printed in Wear, Vol. 34 (1975).

J. Halling, "A Contribution to the Theory of Mechanical Wear," pp. 239-249.

A. A. Reda, R. Bowen, and V. C. Westcott, "Characteristics of Particles Generated at the Interface Between Sliding Steel Surfaces," pp. 261-273.

N. P. Suh and P. Sridharan, "Relationship Between the Coefficient of Friction and the Wear Rate of Metals," pp. 291-299.

which also refers to

J. F. Archard, "Contact and Rubbing of Flat Surfaces," Journal of Applied Physics, Vol. 24 (1953), pp. 981-988.

N. P. Suh, "The Delamination Theory of Wear," Wear, Vol. 25, (1973) pp. 111-124.

Acknowledgement

It is a pleasure to acknowledge the extremely helpful discussions with my colleagues J. R. Rice on the dimensional analysis and B. Budiansky on the general topic of wear. This

research was supported by the Advanced Research Projects Agency
of the Department of Defense under Contract No. DAHC15-71-C-0253
with The University of Michigan.

THE ROLE OF CRYSTAL PLASTICITY IN THE
DELAMINATION THEORY OF WEAR

J. P. Hirth

ABSTRACT

Some implications of crystal plasticity effects with respect to the delamination theory of wear are discussed. Critical experiments to decide the soft versus hard surface question, of interest both with respect to wear and to deformation in general, are suggested. A specific dislocation model for ncp metals under wear conditions is proposed.

THE ROLE OF CRYSTAL PLASTICITY IN THE
DELAMINATION THEORY OF WEAR

J. P. Hirth

I. INTRODUCTION

In considering the delamination theory of wear, certain aspects of the theory, mainly relating to crystal plasticity, appeared to be open to alternative mechanistic interpretation. These are the question of soft versus hard surface layer, which is a controversial topic in even single crystal deformation²⁻⁴, and a model for the low wear rates corrected with certain hcp metals. Models for both phenomena are discussed below, together with suggestions for experiments to test the ideas. Also a brief discussion is presented of the wear equation and several experiments are suggested relative to the phenomenology of the wear process. These comments are relevant to some detailed mechanistic aspects of the delamination theory and not to the phenomenological aspects of the theory¹, for which the evidence is completely convincing.

II. THE SOFT-HARD SURFACE PROBLEM

One part of the delamination theory envisions a soft surface layer in the surface region, and a preliminary rationale¹ of this involved the concept of dislocation removal from this region under the action of image forces. Here, we suggest

that compatibility effects are a more likely cause of a soft surface layer in the $10\mu\text{m}$ scale, and first discuss this in terms of observations on single crystals deformed in tension.²⁻⁴

Fourie² deformed single crystals of copper and then sliced the crystals into wafers parallel to the original surface and measured the flow stress of the wafers. He found a lower flow stress for the surface, on a $\sim 10\mu\text{m}$ to $\sim 100\mu\text{m}$ scale of distance from the surface, relative to the bulk. Mughrabi³ measured dislocation densities as a function of distance from the surface for deformed copper single crystals and, using the empirical relation between flow stress and square-root of dislocation density, found agreement with the Fourie work. Contrariwise, Kramer⁴ has deformed aluminum single crystals, etched off layers ~ 0 to $\sim 10\mu\text{m}$ thick, and measured the flow stress of the remaining crystal. His work indicates surface hardening relative to the bulk. A model which rationalizes both types of behavior follows.

As indicated in Fig. 1, a film free surface should act as a preferential dislocation source^{5,6} and an easy point of egress of dislocations, because of the image force attraction.⁷ At the first grain boundary encountered, however, plastic⁸ and anisotropic elastic⁹ compatibility effects will be present, and after about a three grain penetration into the specimen, the full bulk compatibility constraints will be present. These effects introduce multiple slip at the boundaries as shown, and, as is well known¹⁰, preferential hardening in the multiple

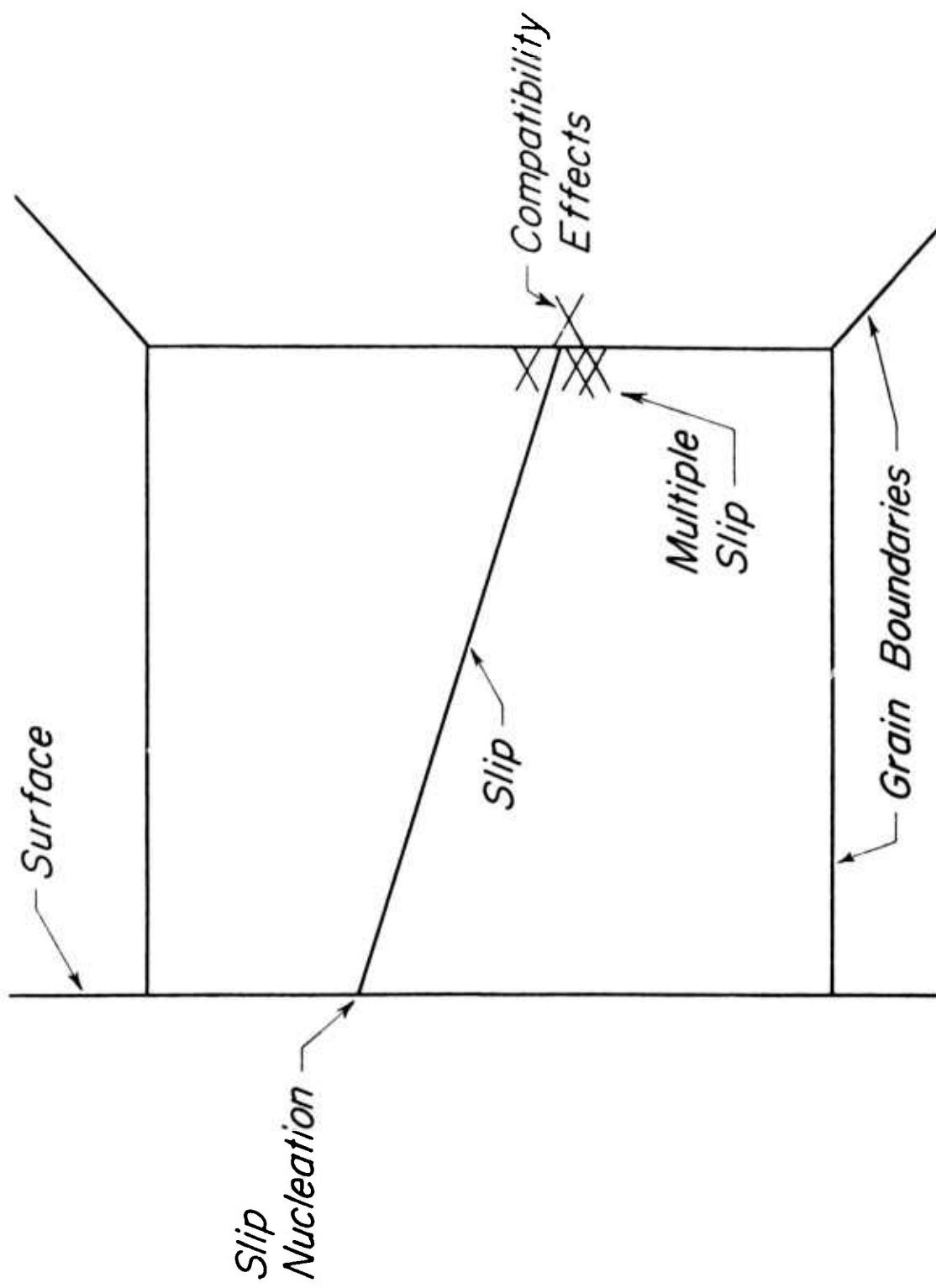


Figure 1. Crosssection of surface showing preferential dislocation nucleation at the surface and dislocation pileups causing multiple slip by combined plastic-elastic compatibility effects at a grain boundary.

slip region spreads from the boundaries. Both this spread of multiple slip and the incubation of multiple slip within the grain interiors are favored by an increase in the tendency for cross-slip of dislocations, which in turn is related to an increase in stacking fault energy. Thus, low stacking fault energy materials should preferentially exhibit soft surface regions relative to high stacking fault energy materials. The scale of this softening will be of the order of 1 to 3 grain diameters ($>10\mu\text{m}$) early in deformation or when the grain diameter is of the order of the stable dislocation cell size later in deformation. However, for plastic strains exceeding ~ 0.1 , a stable cell size tends to form^{11,12} and if it is less than the grain size, it can be the pertinent controlling dimension in Fig. 1.

Empirically,^{11,12} stable cell sizes scale roughly as

$$\lambda \propto 1/\sigma \dot{\varepsilon}^n \varepsilon^m \exp(A/T) \quad (1)$$

where σ is the flow stress, $\dot{\varepsilon}$ is the strain rate, n and m are constants less than unity and T is absolute temperature. As a specific example, for tensile deformation of iron^{11,12} the flow stress-cell size relation at steady state is about

$$\sigma = \mu b / 3\lambda \quad (2)$$

where μ is the shear modulus and b the magnitude of the Burgers vector. If the stacking fault energy is high, the cells tend to be strong barriers resembling grain boundaries

(except for cryogenic temperature deformation or the early stages of high cycle fatigue) whereas if it is low, the boundaries tend to be semi-permeable to dislocations.¹³ Thus, again low stacking fault energy materials tend to have softer surface regions, although the size scale of the region is greatly reduced when cell formation occurs.

A second factor influencing the compatibility effects is the number of available slip systems.¹⁴ With many available slip systems, as with cubic metals, there is more tendency for the transition to multiple slip than with a limited number of slip systems. There are complications in this trend at low temperatures where twinning can occur, with the consequences of obviating the simple correlation.

With surface oxide films present, as in Fig. 2, again the dominant effect is one of compatibility¹⁵, although dislocations will now also experience a repulsive image force. Multiple slip is now generated at the surface which should be preferentially hard relative to the interior.

On a finer scale, image forces play a role in the process of Fig. 1.¹⁵ If dislocations are inclined to the surface, as in Fig. 3, image forces will tend to cause cross-slip and/or climb of a dislocation out of its glide system so that its line is normal to the surface. As an upper limit, these processes will occur over the distance λ of Eqs. (1) and (2). More likely, because of pinning effects and because of the activated nature of the cross slip and climb processes,¹⁴ it will occur over

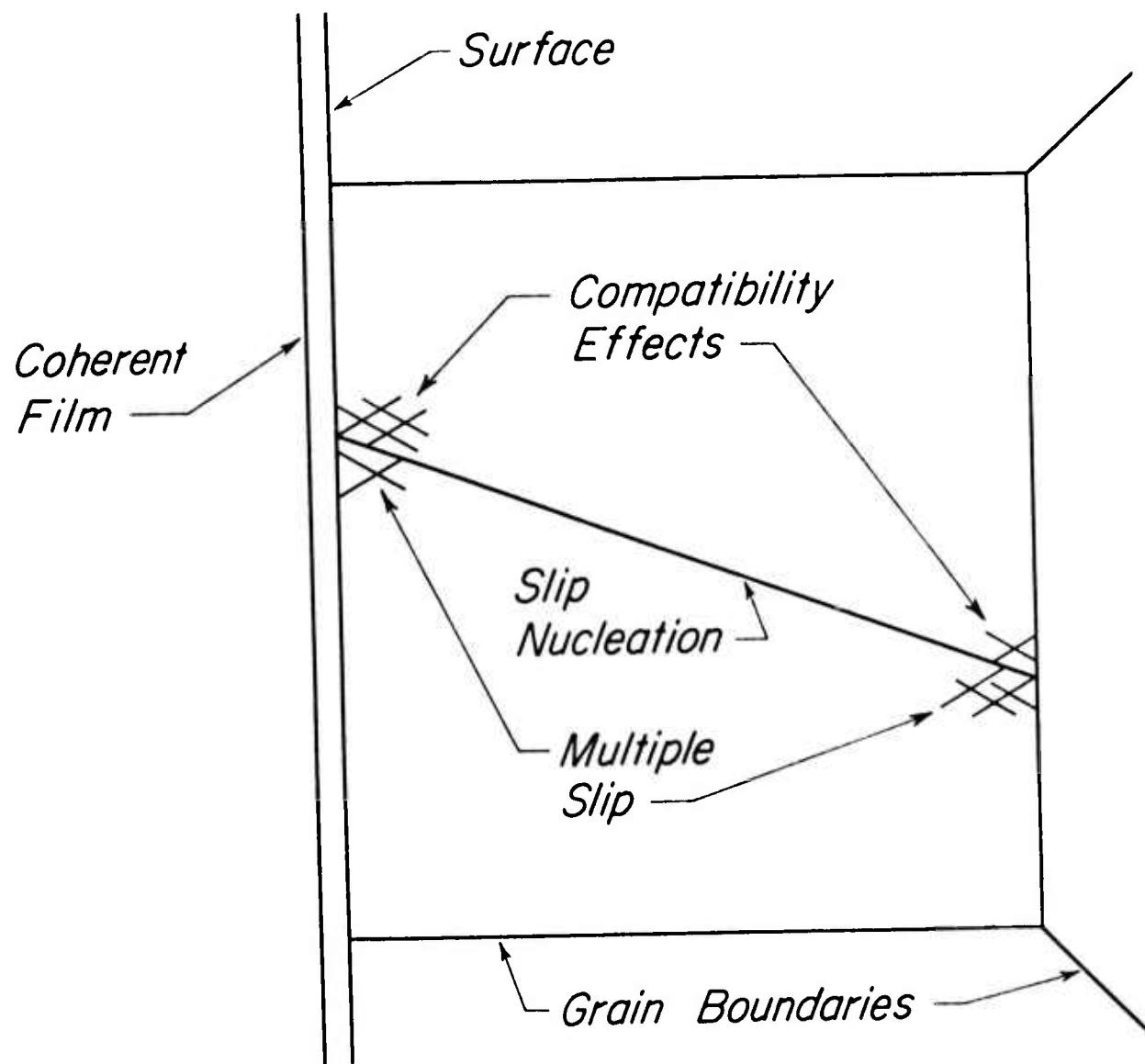


Figure 2. View as in Fig. 1 but with an adherent surface film. Dislocation nucleation is now within the grain with compatibility effects at the surface and at grain boundaries.

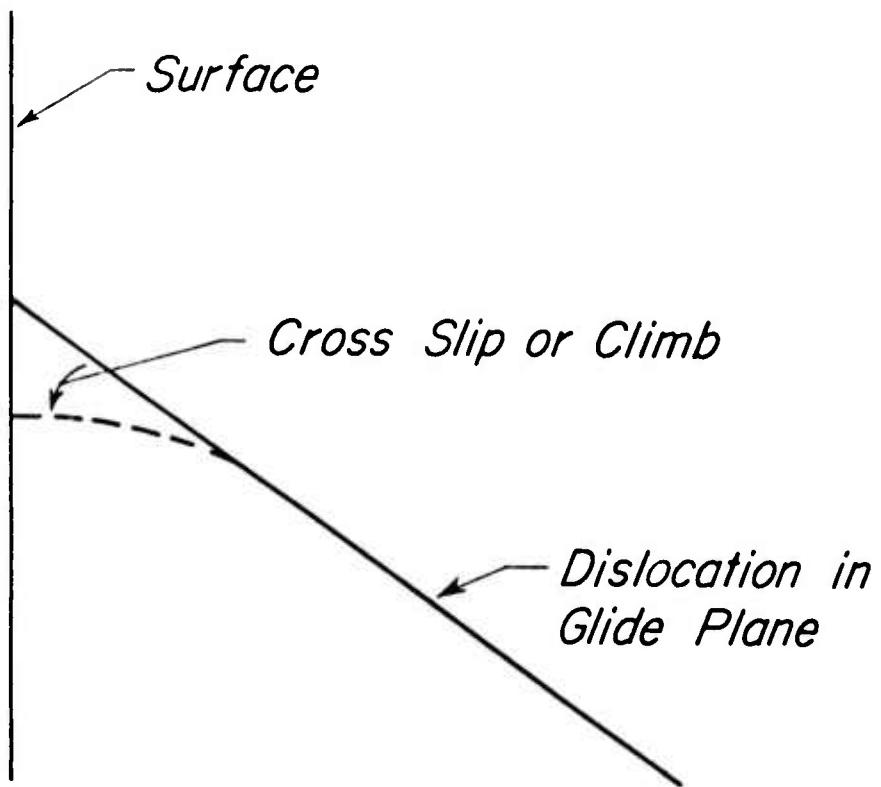


Figure 3. Cross-slip or climb of a dislocation near a free surface under the influence of image forces.

distances a factor of from 3 to 10 smaller, with the latter value applicable to low stacking fault energy materials. This image effect will produce jogs on an atomic scale^{14,15} which will locally produce direct hardening as well as promoting dislocation intersection which would produce surface hardening indirectly. The cross-slip and climb processes would be favored in materials with high stacking fault energy, so again the correlation of surface hardening with this parameter appears.

The consequence of the above discussion are presented in Table 1. Both the fine scale surface hardening observed by Kramer⁶, blocks 7 and 8 in the table, and the coarser scale softening observed by others^{2,3}, blocks 1 and perhaps 2, are rationalized. The seemingly controversial results on surface microhardness¹⁶ as a function of strain, while less directly interpretable, can similarly be rationalized in terms of the table.

Turning to wear, where steady state surface strains¹⁷ are $\epsilon > 15$, we see that it is questionable whether surface softening is still present according to the above discussion. Indeed, as suggested by Budiansky¹⁸, surface softening may not be a requisite for wear to occur by ductile fracture below the surface in the delamination theory, the stress state perhaps being a more critical factor. A CRITICAL EXPERIMENT which would resolve the issue would be to perform wear experiments and near-surface hardness measurements on materials with similar elastic properties but with greatly different stacking

TABLE 1. Summary of surface hardening effects as influenced by stacking fault energy γ and plastic strain ϵ for the grain size scale 1-10 μm and the single dislocation-dislocation cell scale 0.1-1 μm . Relative hardening indicated by A, high hardening greater than or equal to bulk hardening; B, moderate hardening, of the order of bulk hardening; and C, softening relative to bulk. Here γ is shear modulus and b is Burgers vector magnitude.

	A- very near grain boundaries because of compatibility, but in grain interior	(1)	B- compatibility effects at cell boundaries, but boundaries permeable to dislocations	(2)	1-10 μm
		—	—	—	
γ low ($< 5 \times 10^{-3} \text{ J} \mu\text{b}$)	B- hardening because of image effects reduced	(3)	B- Same as (2)	(4)	0.1-1 μm
	—	—	—	—	—
γ high ($> 5 \times 10^{-3} \text{ J} \mu\text{b}$)	A- compatibility effects at grain boundaries	(5)	A- compatibility effects at cell boundaries	(6)	1-10 μm
	A- hardening because of image effects	(7)	A- compatibility effects at cell boundaries and image effects	(8)	0.1-1 μm

fault energies. Sample pairs would be Cu-2%Ge solution versus Cu of commercial purity or with small amounts of Fe in solution. The elastic properties would be the same within about 1%, but the stacking fault energy would be $\sim 10 \text{ mJ/m}^2$ in the former case, $\sim 80 \text{ mJ/m}^2$ in the latter.¹⁹ A similar system, but with somewhat less discrimination, would be¹⁹ OFHC Cu, $\gamma \sim 80 \text{ mJ/m}^2$, and Cu-10%Zn, $\gamma \sim 15 \text{ mJ/m}^2$.

A second experimental area relevant to the wear problem would be further electron transmission microscopy studies of the wear surface layer on the dislocation scale. The phenomenology embodied in Eq. (1) has been developed largely in tensile tests with $0 < \epsilon \lesssim 1$. It may be that for the much larger strains involved in the wear process, the dynamic recrystallization temperature is reduced to near room temperature so that this process too has an influence on surface hardening-softening. (This is definitely the case for high velocity wear²⁰ where the temperature rise at the wear surface is large, $> 200^\circ\text{C}$, but is a moot point for lower velocity wear where the temperature rise is minimal.) In any case, such work would definitively resolve the role of image stresses versus compatibility effects in surface hardening-softening during wear.

III. A CRYSTAL SLIP MODEL FOR WEAR OF hcp METALS

Some hcp metals, those with greater than ideal c/a lattice parameter ratios, are characterized as having anomalously low wear rates and as developing a texture with the

basal plane normal to the wear surface^{1, 21}. There is a possible interpretation of these effects in terms of a crystal plasticity model. The hcp metals with high c/a ratios have very large variations of flow stress with orientation,¹⁴ e.g., a factor of 200 difference between $\langle 11\bar{2}0 \rangle$ (0001) slip and the next less favorably ranked $\langle 11\bar{2}3 \rangle$ (1122) system in cadmium.²² Hence, basal slip is highly favored and the (0001) preferred orientation of Fig. 4 should develop as a consequence of a nonlinear, crystal-plastic, "deck-of-cards" type shear.

Once developed, the structure of Fig. 4a could produce a ratchet-type shear without damage or hardening, hence giving a low wear structure, as indicated in Figs. 4b and 4c. In the deformation step, edge dislocations in small angle grain boundaries glide to produce plastic shear, Fig. 4b. In the recovery step, the edge dislocations glide without net shear to restore the starting structure, Fig. 4c.

A simple test of this idea would be to perform electron transmission microscopy on, say, a cadmium surface as a function of wear to determine whether the necessary stable edge dislocation cell network formed. If the model is valid for bulk Cd material, it is conceivable that it may also apply for the 10^{-7} m coatings of Cd on other materials, which have been found to greatly increase wear resistance¹¹. A simpler test in the latter connection would be to study possible texture development in the coating as a function of sliding contact.

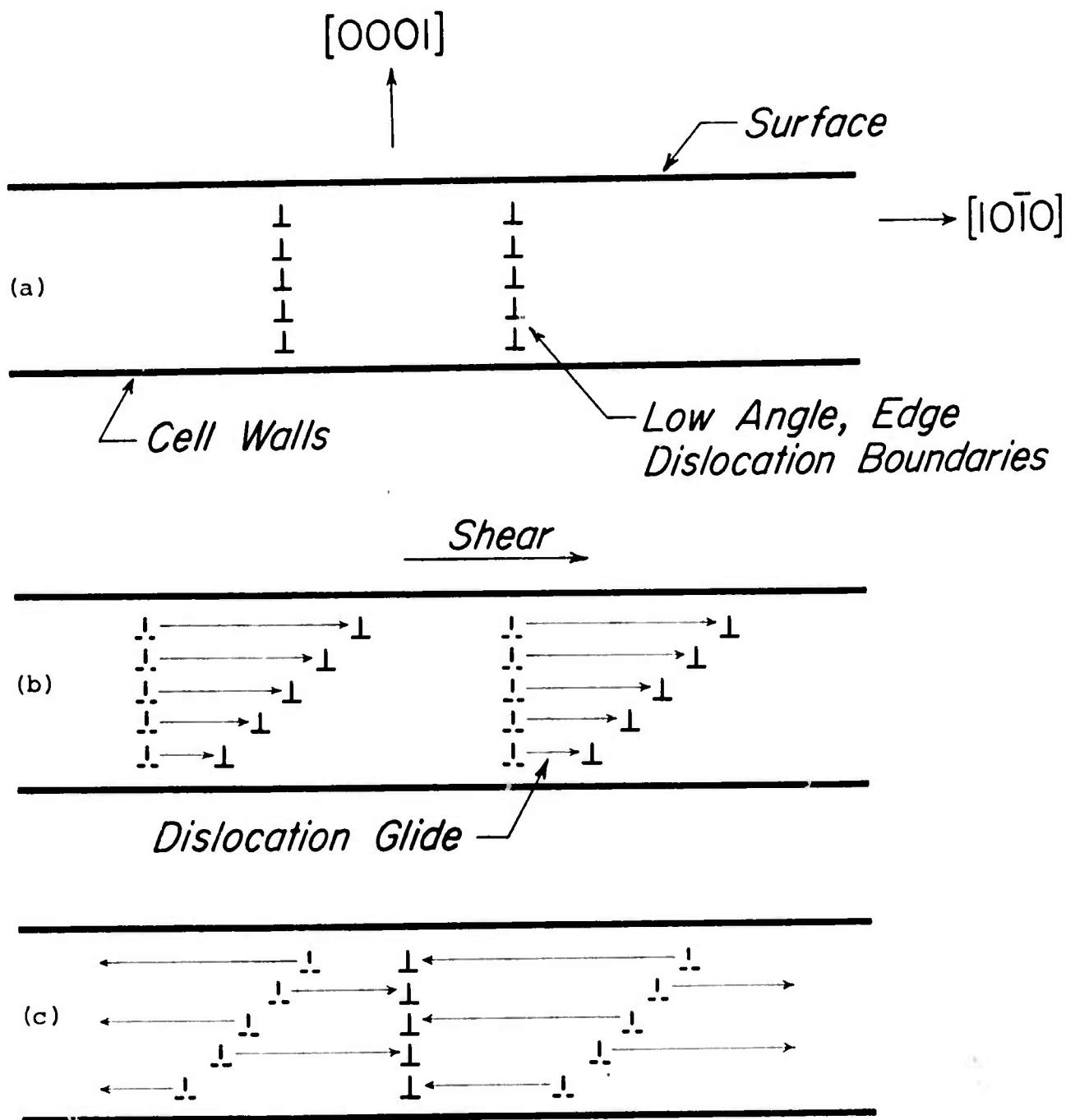


Figure 4. Views of the surface of a cadmium crystal along a $<10\bar{1}0>$ direction and parallel to the (0001) glide plane which is also parallel to the surface. a) Low angle edge dislocation tilt boundaries. b) Glide of the edge dislocations producing shear. c) Glide of the edge dislocations without shear to recover the tilt boundary structure.

IV. SPECIFIC SUGGESTED EXPERIMENTS

First to summarize experiments suggested above:

a. Compare wear rates and surface hardness for materials with nearly the same elastic properties and gross microstructure but with different stacking fault energy (e.g., Cu versus Cu-2%Ge).

b. Study dislocation cell size as a function of distance from the contact surface after wear with a variety of experimental variables but in the range where delamination theory applies.

c. Study surface crystallographic texture and dislocation cell size and nature in worn surfaces of bulk cadmium and for worn surfaces of, say, steel with a 10^{-7} m coating of cadmium.

Some other random experiments are:

d. Perform an experiment in which a slider is very rapidly lifted from a wear surface. Study the wear surface by scanning electron microscopy and optical microscopy to attempt to ascertain the development of wear topography as a function of local distance slid.

e. Compare tungsten and molybdenum base alloy systems to ascertain the degree to which anisotropic elasticity compatibility effects influence wear (both Mo and W are bcc with high stacking fault energy, but tungsten is effectively isotropic while molybdenum is moderately anisotropic in elastic properties).

ACKNOWLEDGEMENT

Helpful discussions with N. P. Suh and the support of this work by the Advanced Research Projects Agency of the Department of Defense Contract No. DAHC15-71-C-0253 with The University of Michigan are gratefully acknowledged.

References

1. N. P. Suh, Wear 25, 11 (1973); with S. Jahanmir and E. P. Abrahamson, Wear 28, 235 (1974); ARPA Report AD/A-003 583, September 1974.
2. J. T. Fourie, Phil. Mag. 17, 735 (1968).
3. H. Mughrabi, Phys. Stat. Sol. (6) 44, 391 (1971).
4. I. R. Kramer, Trans. Met. Soc. AIME 233, 1462 (1968).
5. F. C. Frank, in "Symposium on Plastic Deformation of Crystalline Solids," Carnegie Institute of Technology, Pittsburgh, 1950, p. 89.
6. J. P. Hirth, in "Relation between Structure and Strength in Metals and Alloys," H. M. Stationery Office, London (1963) p. 218.
7. A. K. Head, Phil. Mag. 44, 92 (1953).
8. J. D. Livingston and B. Chalmers, Acta Met. 5, 322 (1957); 6, 216 (1958).
9. R. E. Hook and J. P. Hirth, Acta Met. 15, 535 (1967).
10. U. F. Kocks, Met. Trans. 1, 1121 (1970).
11. A. S. Keh and S. Weissman, in "Electron Microscopy and Strength of Crystals," eds. G. Thomas and J. Washburn, Interscience, 1963, p. 131.
12. G. Langford and M. Cohen, Trans. ASM 62, 623 (1969).
13. D. J. Boyd and J. D. Embury, in Proc. 5th Int. Conf. Materials, Berkeley, California, 1972.
14. J. P. Hirth and J. Lothe, Theory of Dislocations, McGraw-Hill, 1968.
15. J. P. Hirth, Met. Trans. 3, 3047 (1972).
16. N. Gane and J. M. Cox, Phil. Mag. 22, 881 (1970).
17. J. H. Dautzenberg and J. H. Zaat, Wear 23, 9 (1973).
18. B. Budiansky, Report of the ARPA Materials Research Council, Univ. of Michigan, Ann Arbor, July 1975.

19. P. C. J. Gallagher, Met. Trans. 1, 2429 (1970).
20. R. C. Bill and D. W. Wisander, NASA TN D 7840, Dec. 1974.
21. E. Rabinowicz, Friction and Wear of Materials, Wiley, 1965.
22. P. B. Price, in "Electron Microscopy and Strength of Crystals," Interscience, 1963, p. 41.

WEAR, POLISHING AND THE STUDY OF
SURFACE LAYERS BY X-RAY DIFFRACTION

A. Bienenstock

ABSTRACT

In this note, the determination of surface layer characterization parameters, crystallite size and average strain, through the use of X-ray diffraction with wavelengths just below absorption edges is discussed. The note begins with calculations of penetration depths at these wavelengths and an analysis of the contribution of surface layers to the measured diffracted intensities. After some comments on experimental techniques, applications of high absorption diffraction techniques to the study of worn metal surfaces and to quality control of molybdenum high power infrared laser mirror surfaces are proposed.

WEAR, POLISHING AND THE STUDY OF SURFACE LAYERS
BY X-RAY DIFFRACTION

A. Bienenstock

I. Introduction

In recent years, MRC has studied problems such as wear of metals and the polishing of mirror surfaces where structural characterization of 1 to 10 micron thick surface layers would be highly desirable. In the problems involved, even the determination of the average crystallite size and strain or some knowledge of texturing would be quite advantageous. Unfortunately, nondestructive techniques for this purpose have not been proposed. The layer is too thin for normal X-ray diffraction experiments and is too thick for electron diffraction. Consequently, the latter technique is occasionally combined with successive surface etches to obtain size and strain as a function of depth. Alternatively, sections cut normal to the surface are studied in the electron microscope.

In this note the determination of surface layer characterization parameters, crystallite size and average strain, through the use of X-ray diffraction with wavelengths just below absorption edges is discussed. The note begins with calculations of penetration depths at these wavelengths and an analysis of the contribution of surface layers to the measured

diffraction intensities. After some comments on experimental techniques, applications of high absorption diffraction techniques to the study of worn metal surfaces and to quality control of molybdenum high power infrared laser mirror surfaces are proposed.

II. X-ray Penetration Depths

In this section it is shown that, through the selection of appropriate wavelengths, the X-ray penetration depth can be made as small as 0.6 to 4 microns for a number of metals. As a result, a normal X-ray diffractometer line broadening study employing such wavelengths would sample only such a depth.

Here, the penetration depth, t , is defined as the thickness of material which transmits $1/e$ of a normally incident X-ray beam. That is,

$$t = 1/\mu , \quad (1)$$

where μ is the linear absorption coefficient. This is a useful parameter since, in a typical surface reflection diffraction experiment, both the incident and diffracted beam are absorbed by the material through which they pass. As a result, material lying deeper than the $1/e$ depth makes a small contribution to the measured diffracted intensity. The actual contribution to the scattering, as a function of depth, is discussed in Section III of this note.

Calculated penetration depths of a number of X-ray wavelengths in Fe, Cu and Mo are shown in Table 1. The wavelengths correspond to K_{α_1} radiation from practical X-ray tube anodes. The calculations follow in the normal manner¹ from tabulated values¹ of the mass absorption coefficients of the elements and the metallic densities. It is readily observed that the penetration depth in Fe can be varied from 4.1 to 64.1 microns through appropriate choice of radiation. The 4.1 micron depth corresponds to the use of Cu K_{α} radiation, which is normally avoided in X-ray diffraction because of the fluorescence problem. As is indicated in Section IV, however, this problem may be overcome easily with modern detecting systems.

Similarly, the penetration depth in Cu can be varied from 4.0 to 42.4 microns while that in Mo can be varied from 2.0 to 77.6 microns. Penetration depths in the range of 0.60 to 0.76 microns can be obtained through the use of Ti K_{α} radiation with the metals Au, Ir, Pt, Rh, and W. Hence, it is possible, through proper choice of X-ray tube anode, to average over surface layers of varying depths down to those of interest in wear experiments.

III. The Actual Depth of Diffraction

In this section the actual depths which are contributing to a measured diffractometer peak are analyzed. For simplicity, we consider the special case where, within the

penetration depth, the sample is homogeneous. It can easily be shown using a slight generalization of the analysis of Klug and Alexander², that the applicable equation is:

$$F = [1 - \exp(-2x\mu/\sin\theta)] = [1 - \exp(-2x/t \sin\theta)] . \quad (2)$$

Here, F is the fraction of the measured intensity which is contributed by the layer of depth x from the surface of a material at a Bragg angle θ . For fixed t , the greatest depth of diffraction is obtained at $\theta = 90^\circ$, where 75% of the diffracted intensity comes from a layer $0.7t$ thick. At $\theta = 45^\circ$, that fraction comes from a layer $0.5t$ thick.

IV. Comments on Experimental Methods

Some of the combinations of X-ray wavelengths and scattering materials discussed here are those traditionally avoided in X-ray diffraction. The closeness of the photon energy to an absorption edge of the scattering material implies the production of considerable fluorescent radiation. Fortunately, detector-pulse height analyzer systems are now readily available which can be set so that none of the fluorescent radiation is counted. Hence, this consideration is no longer operative.

The experiments proposed in Section V of this note could be performed with a normal X-ray generator and a counter-pulse height analyzer system of good (300-600 eV.) resolution. No crystal monochromatization would be required.

The use of synchrotron radiation for this type of experiment has both advantages and disadvantages. The major advantages are tunability of the incident wavelength and higher intensity. The tunability would allow one to employ a wavelength considerably closer to the absorption edge and to obtain, therefore, a larger absorption coefficient and a smaller penetration depth. For example, the penetration depth in Fe could be reduced from the 4.1 microns associated with Cu K_{α} radiation to approximately 3 microns by such tuning.

The high intensities become important when long wavelengths, like that of Ti K_{α} , must be employed. High absorption of the X-rays by the X-ray tube window means that the intensities incident on the sample are low, even when the tube is run at high wattage. Since incident intensities, at this wavelength, are higher by a factor of 10^5 to 10^6 at a facility like the Stanford Synchrotron Radiation Project, it would appear easier to perform the experiment there.

A disadvantage of synchrotron radiation for this purpose is, however, that pulse height analyzers cannot be employed to eliminate fluorescent radiation. This is because the X-rays reach the sample in very short duration pulses of extremely high intensity. As a result, it would probably be necessary to employ both incident and diffracted beam monochromators. The incident beam monochromator would be necessary to eliminate a high flux of inelastically scattered X-rays, coming from incident radiation of higher photon energy, whose

energies would be within the band-pass of the diffracted beam monochromator and which would, therefore, form a source of undesirable background. If the synchrotron energy is sufficiently high, it would also be necessary to employ filters to eliminate λ/n components of the beam which both monochromators would pass. As a result of these disadvantages, it would be my inclination to perform experiments first with standard equipment, unless the tunability feature is particularly important for the examination of some physical phenomenon.

v. Some Applications of High Absorption Diffractometry
Suh³ has proposed the existence of a "soft" layer in metal surfaces which have undergone significant benign wear. This layer may range from 1 to 10 microns in thickness. Below this layer is a region which is heavily cold worked. The existence of such a layer should be determinable through application of the Warren and Averbach⁴ line broadening techniques. In this method, diffraction line shapes are measured carefully and then Fourier analyzed. By comparing Fourier components from different orders of a reflection, it is possible to separate crystallite size and strain effects. By performing this separation with diffraction peaks obtained from a radiation whose penetration depth is of the order of the surface layer and another whose penetration depth is twice that, the existence of the layer should be determinable. This approach is particularly applicable to iron and

high iron content alloys, since $Cu K_{\alpha}$ radiation is employed for the studies. The short wavelength, relative to others discussed in this note, allows for the measurement of the shapes of a number of diffraction peaks.

This technique is not available, however, for those materials requiring long (e.g., $Ti K_{\alpha}$) wavelengths to obtain the small penetration depth. Here, only 2 or 3 reflections can be obtained, so that the separation of size and strain effects cannot be achieved. In these cases, one could only use total broadening as a crude measure of combined size and strain effects. Such measurements have, however, occasionally been quite informative.

Indeed, one possible DoD use for them is for quality control of high power infrared laser mirrors made of polished molybdenum. It has been suggested⁵ that the most important factor limiting the reflectivities and the threshold breakdowns of these mirrors is small crystallite sizes within the skin depth ($\sim 1000\text{\AA}$). Equation (2) indicates that approximately 31% of the Bragg (110) peak scattering of $Ag L_{\alpha_1}$ radiation would come from material within the skin depth. By correlating line widths with reflectivity and breakdown threshold, it should be possible to establish criteria for acceptable mirrors. The virtue of this technique is that it is extremely easy and rapid to execute. As a result, every mirror could be checked prior to use.

It should be noted, in this context, that the development of soft X-ray lasers in this wavelength region is now anticipated. They would make an ideal source of radiation for the procedure suggested in the previous paragraph.

Finally, the method could also be used to measure surface layer texturing effects produced by a variety of processes. Here, one would compare the intensities of, say, $(hk0)$ with $(00l)$ reflections with those obtained from a randomly oriented polycrystalline array.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHIC15-71-C-0253 with The University of Michigan.

TABLE 1. Penetration Depths

Material	1/e Penetration Depth in Microns for Various K_{α} Radiations - in Å						
	Ag-0.5608	Mo-0.7107	Cu-1.5418	Fe-1.9373	Mn-2.1031	Cr-2.2909	Ti-2.7496
Fe	64.1	32.8	4.1	19.0	15.0	11.7	6.9
Cu	42.4	22.0	21.2	10.9	8.6	6.7	4.0
Mo	22.0	77.6	8.8	4.8	3.8	3.1	2.0

References

1. International Tables for X-Ray Crystallography, Vol. III, (Kynoch Press, Birmingham, England, 1962). Sec. 3.2.
2. H. P. Klug and L. E. Alexander, X-Ray Diffraction Procedures for Polycrystalline and Amorphous Materials, (Wiley, New York, 1954). p. 376.
3. N. P. Suh, S. Jahanmir and E. P. Abrahamson, The Delamination Theory of Wear, Report AD/A-003 583, available from NTIS.
4. B. E. Warren and G. L. Averbach, J. Appl. Phys. 23, 497 (1952).
5. A. Bienenstock, in Preliminary Reports, Memoranda and Technical Notes of the Materials Research Council Summer Conference, July, 1974. Vol. I, p. 42.

SURFACE MECHANICS AND WEAR

F. F. Ling

ABSTRACT

The role surface mechanics might play in the study of wear is illustrated through several direct experiences. Some of the examples are chosen to illustrate how mechanics principles are used in a definitive fashion. At least one example was chosen to illustrate that there is very little mechanics can do for the wear problem concerned.

In the problems examined, often something is interposed at and around the interface of normal working surfaces. Since all materials do wear, the direct consequence is that, more often than not, the subject of wear study has been transposed to those of coatings, composites and compounds.

SURFACE MECHANICS AND WEAR

F. F. Ling

Surface mechanics¹ marshalls the principles of mechanics, especially continuum mechanics, for the resolution of questions arising from surface, surface layer and interface phenomena. The original definition, laid down some ten years ago, included certain analytical techniques which were particularly suitable for surface work; this point, however, does not seem important here. How relevant is surface mechanics to wear studies? In what way is it most useful? In what way is it useless? The user will have to be the ultimate judge. What will be done here are some examples which were experienced at the Institute for Wear Control Research at RPI.

Strange as it may seem, the name wear control was introduced not too long ago to emphasize some positive steps that characterize the Institute's activity towards mechanical reliability and safety, energy conservation, material conservation and substitutability. This means, often, something is interposed at and around the interface of normal wearing surfaces. Since all materials do wear², the direct consequence is that, more often than not, the subjects of wear study have been transposed to those of coatings, composites and compounds. This fact may add a new dimension to this workshop in its inquiry

into the scope of wear of material systems.

As an example of wear control, there is the ongoing project of coatings for ice breakers so that they can perform their function from which they have been prevented under certain severe conditions of ice. From the fundamental point of view, the desired qualities are low surface adhesion and good bonding to the substrate at the same time, low friction and low wear. The great success to date is a hybrid coating on the U.S. Coast Guard Cutter Raritan. It has a wear life of one year as compared with other coatings which have lives measured by a few days under severe operating conditions. In this example, mechanics principles were used to obtain bounds on the stresses which prevail at the contact areas.

In another situation mechanics principles were used intensively in the case of fretting wear^{3,4}. Fretting wear is one which occurs under concentrated contacts while the loaded surfaces undergo small amplitude oscillatory motion. For this case, contact deformation and stresses were evaluated so as to provide information on microslips, relative displacements between surface elements on two sides of the contacting bodies. Since wear has been experimentally related to microslip, an optimization procedure was invoked. This resulted in a modified gross geometry which would yield minimum wear. Figure 1 shows such a modified geometry together with spherical surfaces in contact. Moreover, the ideal geometry resulted in a 60% decrease in the so-called slip region and the maximum microslip

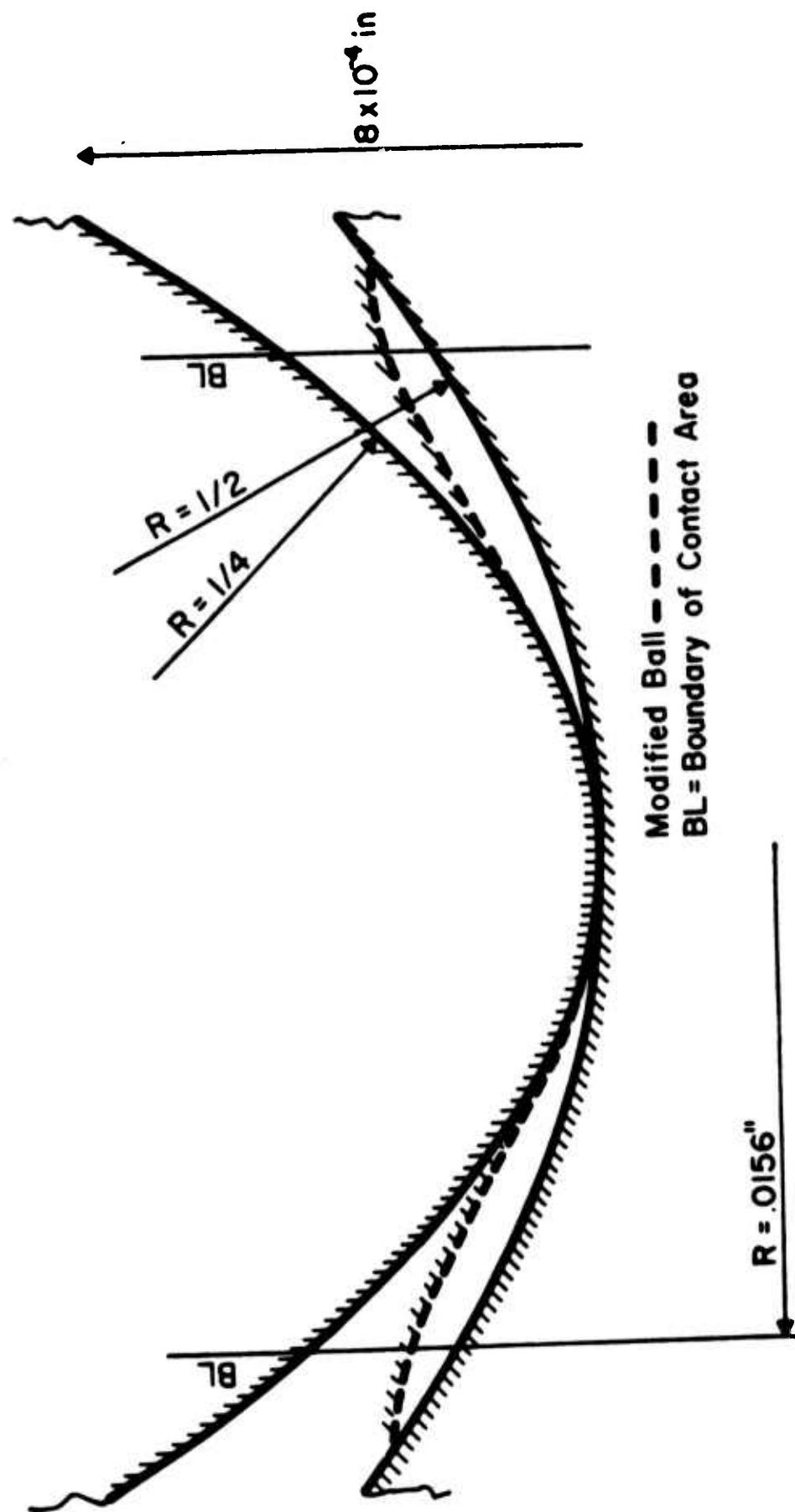


Figure 1.

has been reduced by 90%. Such a surface has been made and tested; the reduction in wear has been observed.

Next, mechanics principles have been applied to the inception of wear of cartilage of human joints. Figure 2 shows a micrograph (1000X) depicting the formation of microcraters which are the precursors of wear degradation. Figure 3 shows a model of such a microcrater along with a microfissure. Both models have been analyzed by principles of mechanics; in this case the material is one of layered composites with the surface layer being poroelastic.

In the pages ahead, three facets of a project which shows how mechanics is helpful or helpless in understanding wear and/or wear control are given. The problem area is that of aircraft brake wear. Figure 4 shows a section through an assemblage of discs and brake pads. Figure 5 shows a pad which has worn, having gone through the stages of cracking, spall formation and wear deterioration.

I. Thermal Stress/Wear Analysis

Since brake contact is a high temperature contact, a thermal stress/wear analysis was carried out⁵. Table I shows some of the concerns about the adequate operation of the brake.

Figure 6 shows a section of an axisymmetric ring shaped brake pair which was studied. This analysis included heat conduction, interface apportioning of heat between the bodies in contact, thermal stress and a wear criterion shown below:



Figure 2.



Figure 3.

SEE FIGURE 6

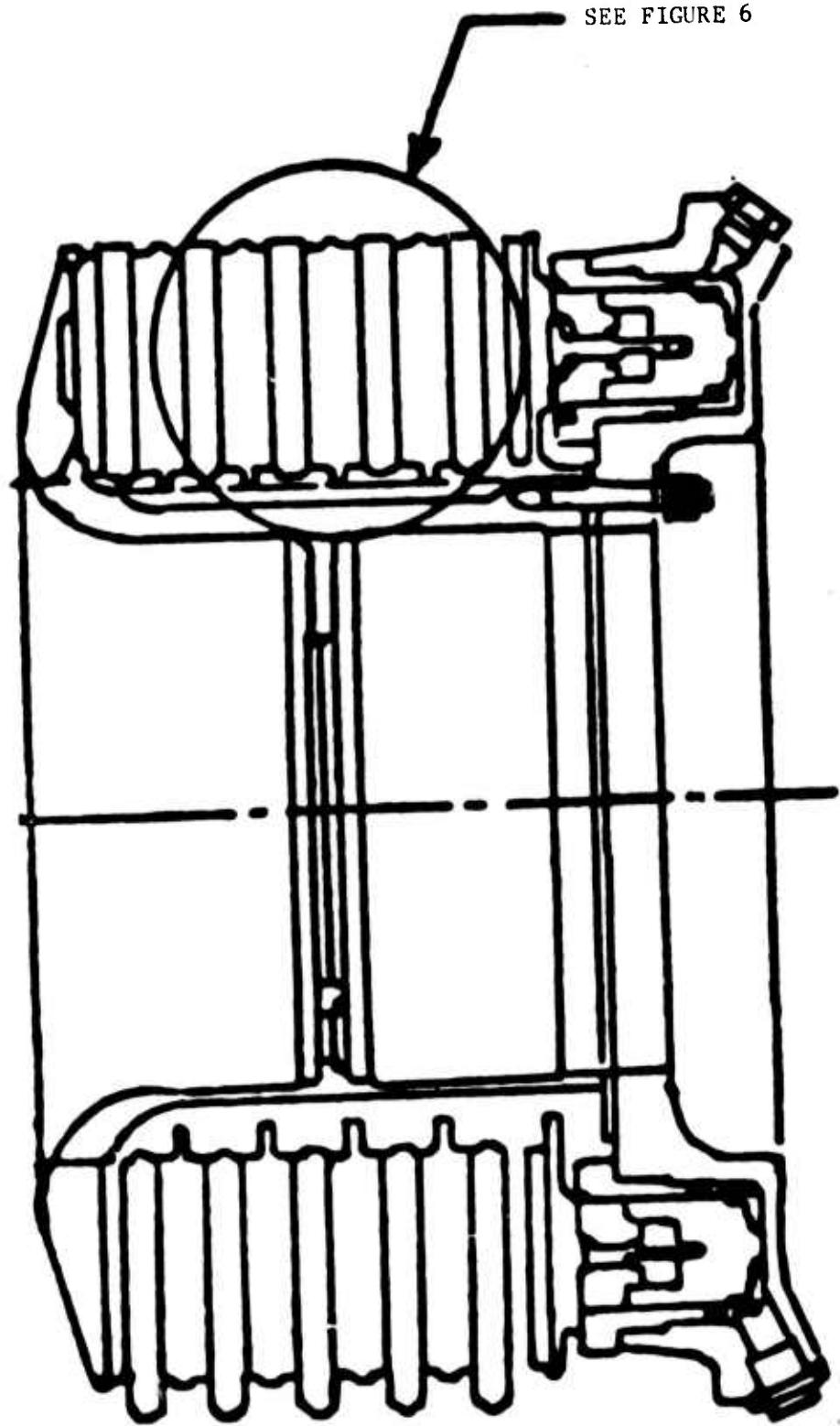


Figure 4.

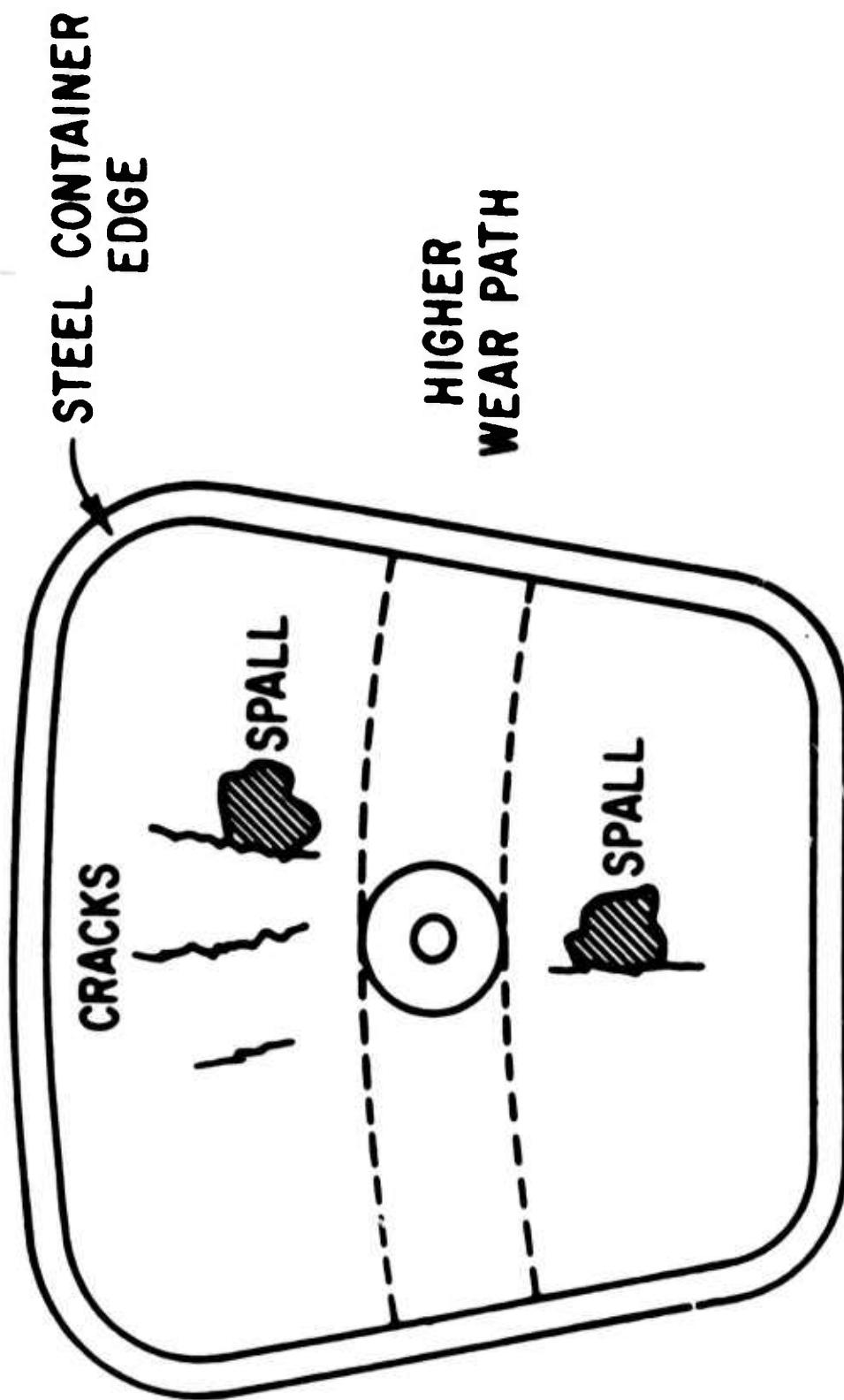


Figure 5.

TABLE I

Adequate strength at operating temperature
Minimum oxidation at operating temperature
Reasonable cost
Non-galling (no transfer)
Constant friction over operating conditions
Minimum wear

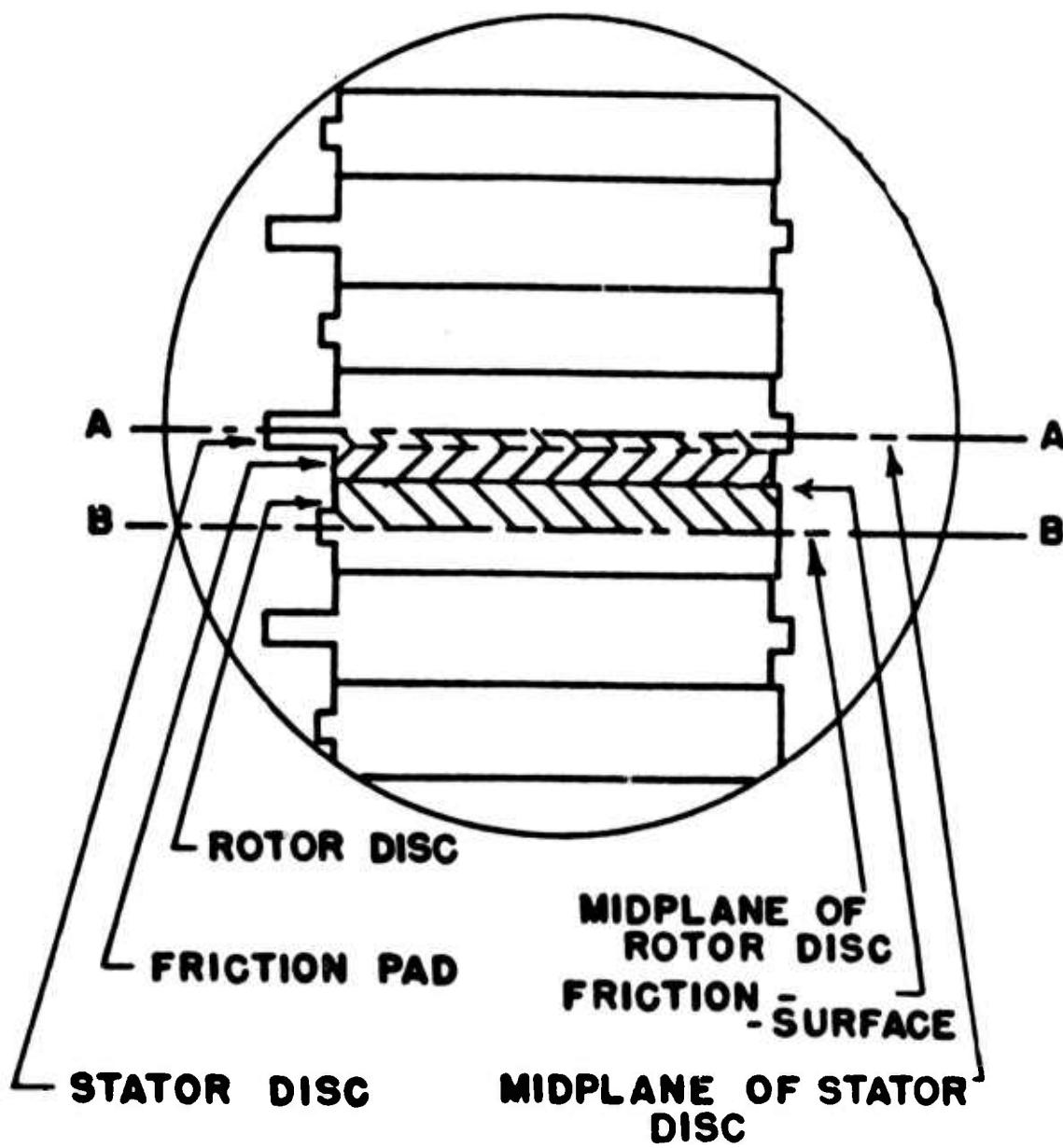


Figure 6.

$$J_2 = W ,$$

where J_2 = second invariant of the stress deviator tensor,

$$W = \frac{1}{3} \sigma_y^2 ,$$

$$\sigma_y = \sigma_y(T) ,$$

σ_y being the yield strength in tension and T is the temperature.

The results are shown in the next three figures. Figure 7 shows how the surfaces are deformed at a given time compared with the undeformed state. Figure 8 shows a sequence of events with respect to time: contact length along the radius of the disc. The dots show when and where discrete wear events have taken place. In other words, the contact area(s), in the form of annulus (or annuli), shifts (or shift) as a consequence of thermal distortion and wear. Figure 9 shows how this theory compares with the constant-contact theory and experimental measurement.

II. Simple Thermal Analysis as a Design Guide

Along with the above, a simple analysis was made to provide the parametric dependence of surface temperature. To keep the surface temperature low, it was found that the desirable quantities are: High thermal conductivity, high specific heat, high density, low thermal expansion, and low elastic modulus. Figure 10 shows the results through the dimensionless parameter $C^* \equiv (dt/l^2)^{1/2}$, where d is the thermal diffusivity,

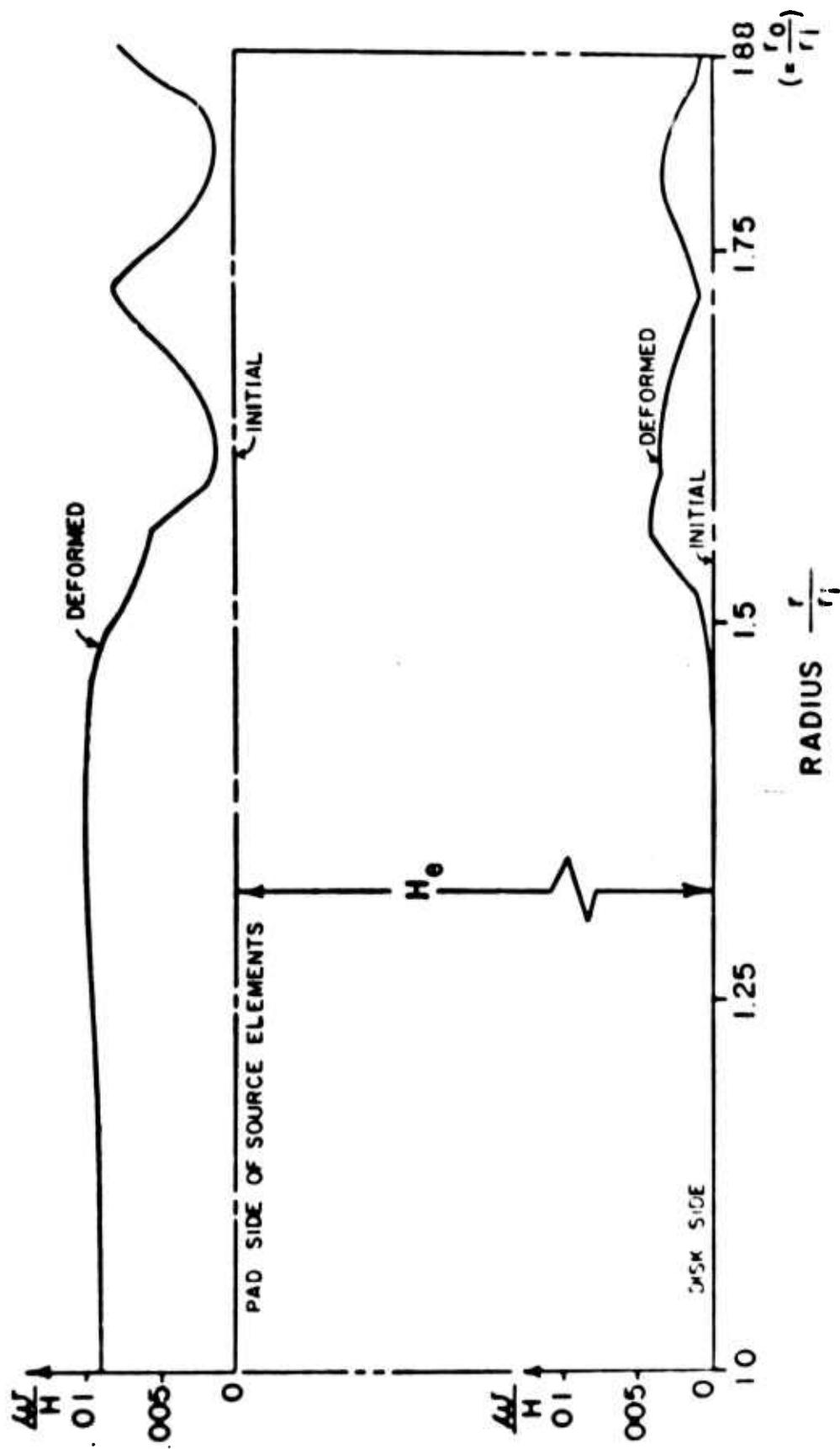


Figure 7.

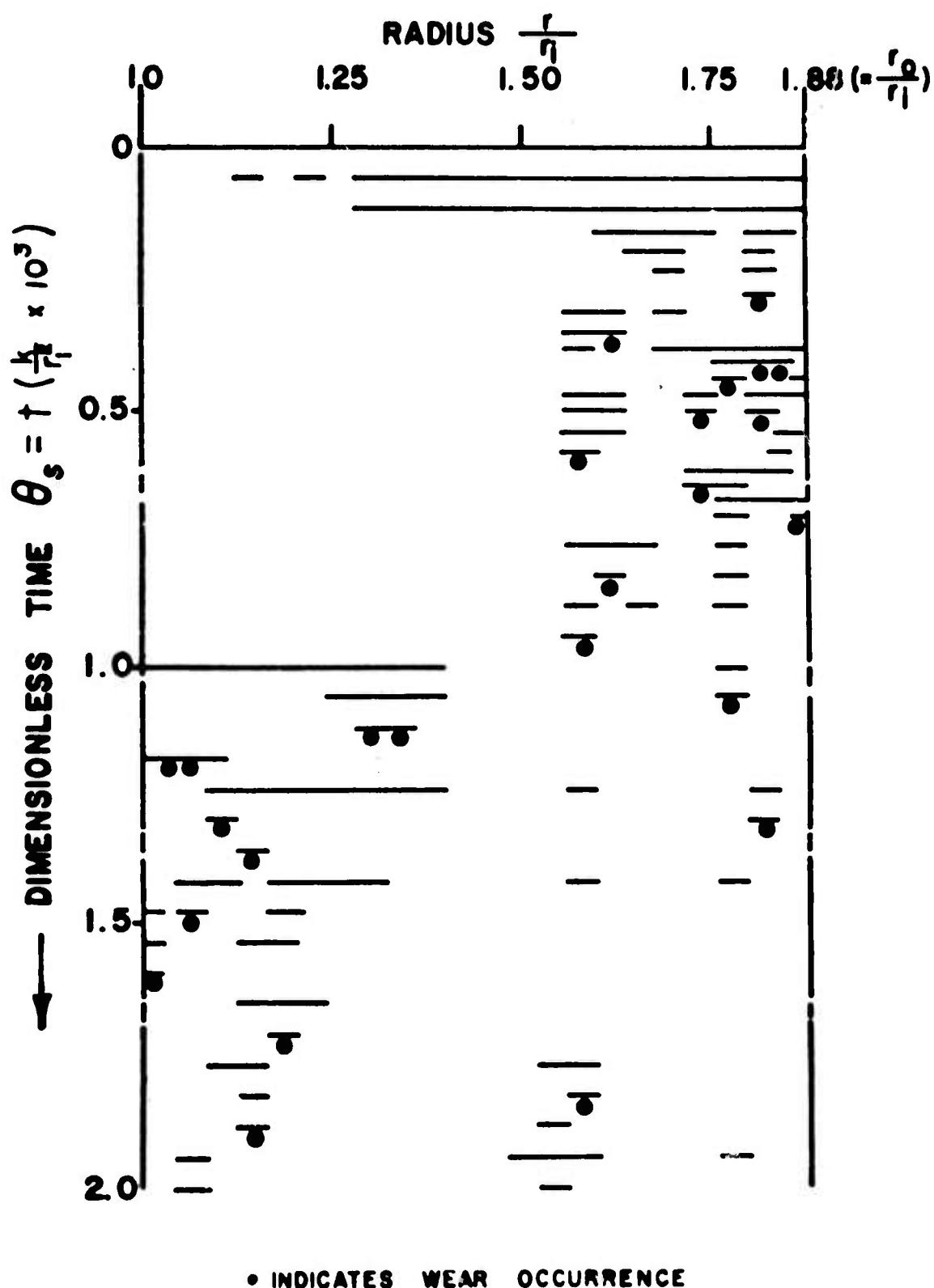


Figure 8.

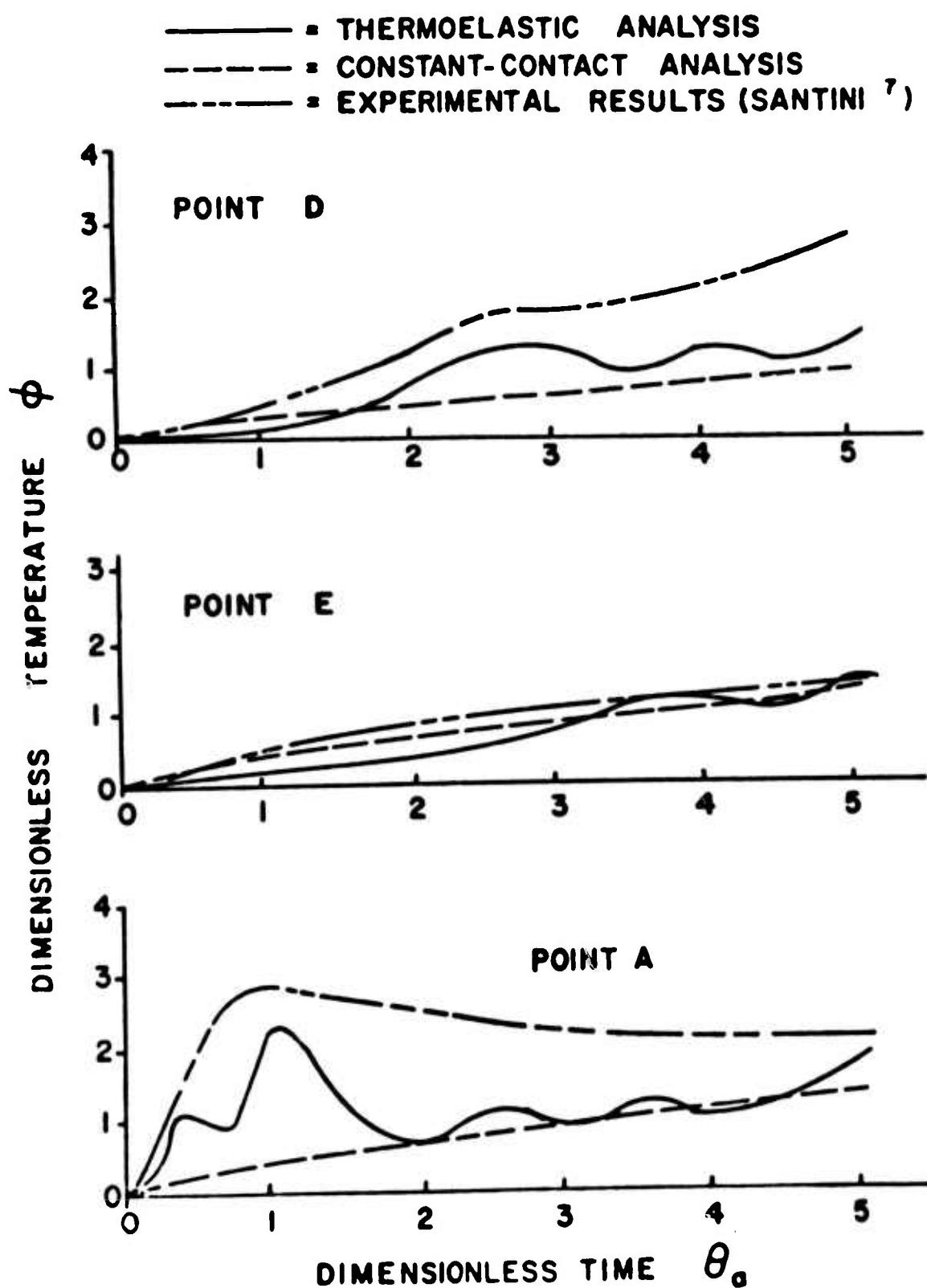


Figure 9.

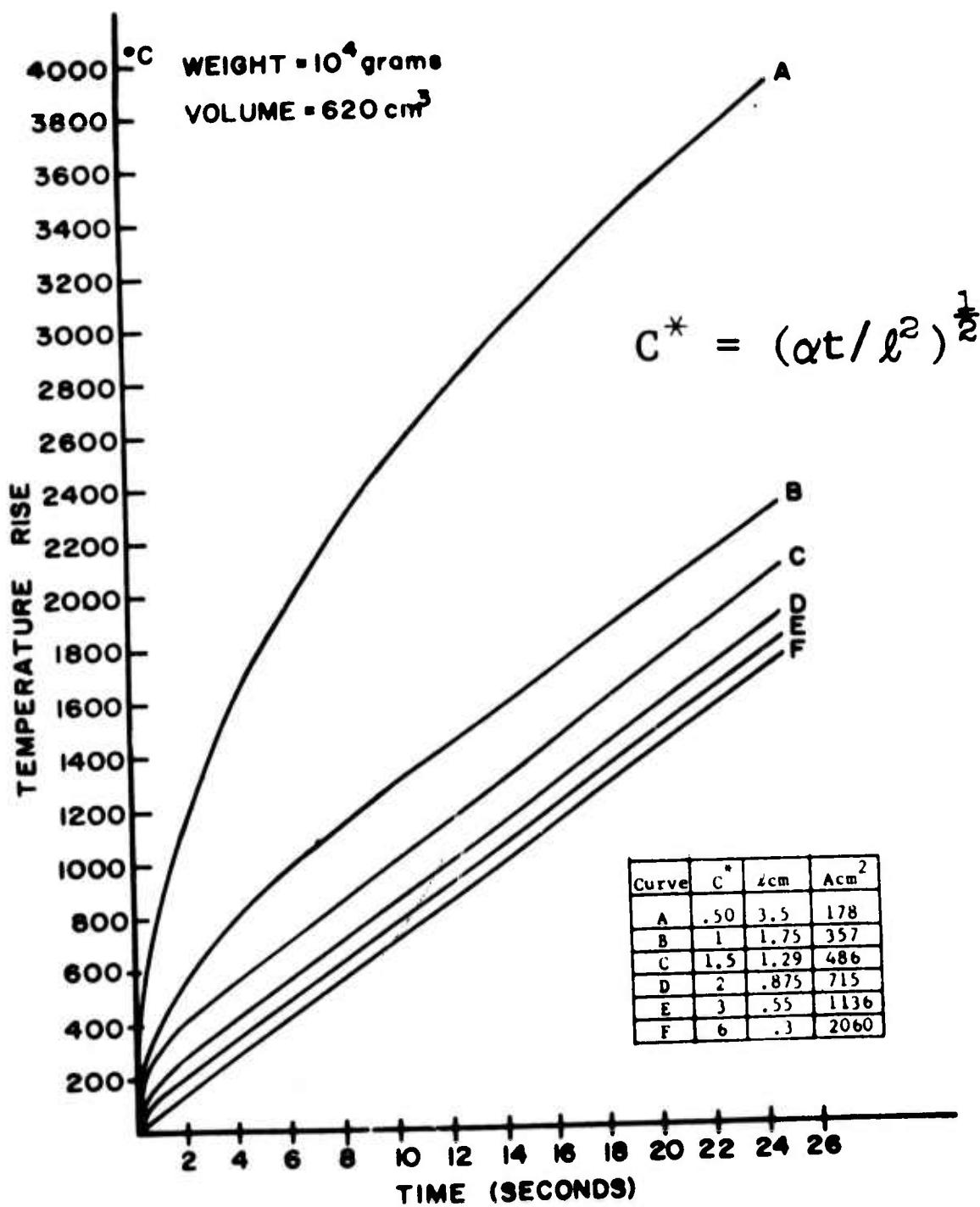


Figure 10.

t is the time, l is the half thickness of the brake disc. In fact $C^* = 1$ is a good rule-of-thumb⁶. That is, for the curves shown, each is concerned with the same volume of brake material. Surface temperature decreases as C^* increases. But the pay-off beyond $C^* = 1$ drastically diminishes and the penalty is on the total area required. Also implied here is the fact that wear affects surface temperature, see Figure 11. For a disc thickness of 1.2 cm ($l=.6$) and an area of 800 cm^2 , which is approximately the actual brake disc of a particular aircraft brake assembly, the temperature rise is 66°K . A wear of .2 cm gives an added temperature of 140K . This added temperature may be significant if the material will be damaged or the wear rate is significantly increased as the result of the temperature rise.

III. Composite Brake Material

Figure 12 shows the result achieved by using composites. In this example, MoS_2 and nickel are used with suitable abrasive. The optimal ratio of MoS_2 to nickel is shown. In this case surface mechanics was only indirectly helpful in arriving at this choice of matrix.

Finally, mechanics has been most useful in the case of requirements on coatings for very high speed application. Since the incipient failure of such films is tied to their fatigue failure and their detachment as spalls, the nature of shear stress at the film-substrate interface is important. It

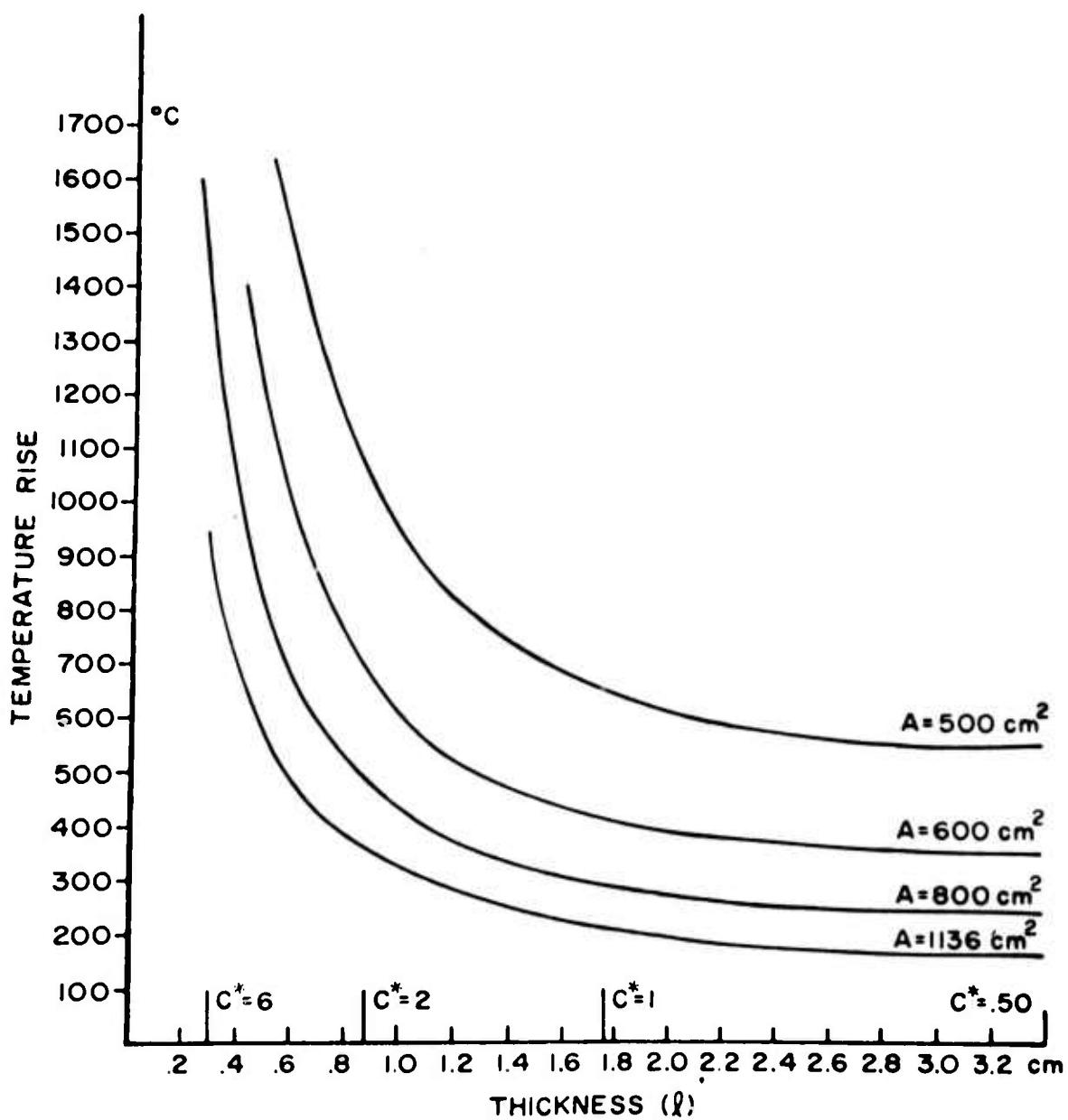


Figure 11.

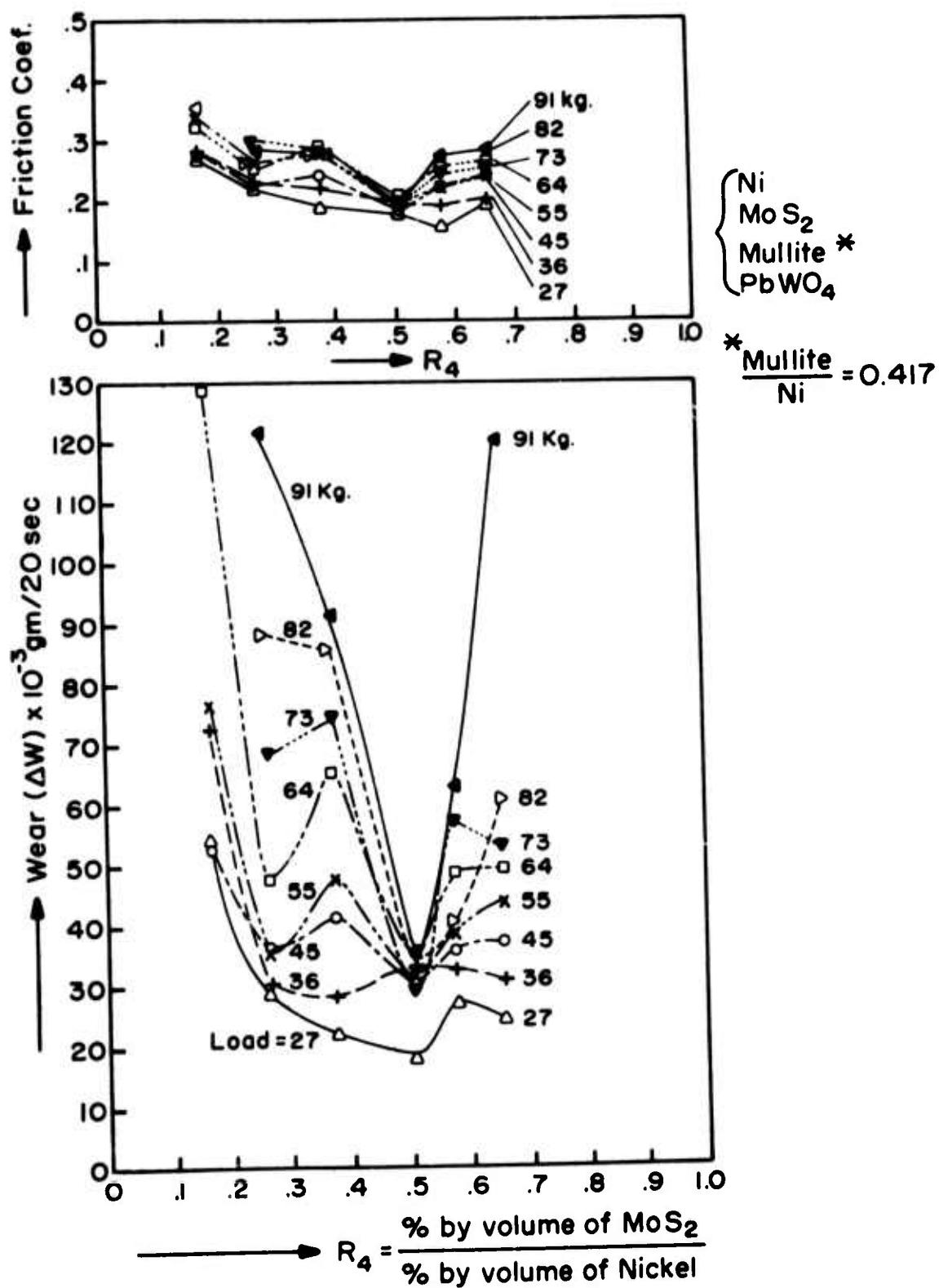


Figure 12.

has been shown that the number of shear reversals in subsonic operation are different from that for transonic and supersonic regimes. Here, sonic speed is meant the speed of propagation of shear waves in the film material and the transonic regime is bracketed between this speed and that of propagation of dilatational waves in the film material. The main results include the finding that there are several stress reversals in the transonic regime as compared to one for the sonic regime⁷.

References

1. Surface Mechanics, F. F. Ling, John Wiley & Sons, New York 1973.
2. Boundary Lubrication - An Appraisal of World Literature, F. F. Ling, E. E. Klaus and R. S. Fein, editors, Americal Society of Mechanical Engineers, 1969.
3. "Microslips Between Contacting Paraboloids," V. C. Mow, P. L. Chow and F. F. Ling, Journal of Applied Mechanics, 34, 321-328 (1967).
4. Fretting in Aircraft Systems, F. F. Ling, editor, North Atlantic Treaty Organization AGARD-CP-161 (1975).
5. "A Thermal, Thermoelastic, and Wear Simulation of a High-Energy Sliding Contact Problem," F. E. Kennedy, Jr. and F. F. Ling, Journal of Lubrication Technology, Transactions of the American Society of Mechanical Engineers, 96, 497-507 (1974).
6. "On the Principles of Aircraft Brakes: The Application of Surface Mechanics," F. F. Ling, Plenary Session, All Union Conference on the Physico-Chemical Mechanics of Contact Interaction and Fretting Corrosion, Kiev, USSR, June 26, 1973. Also appeared in the proceedings.
7. "Shear Stresses in a Layered Elastic System under a Moving Load," Y. C. Hsu and F. F. Ling, Recent Advances in Engineering Science, II, Gordon and Breach, New York, 323-351 (1965).

COMMENTS ON THE CHARGE TO THE ARPA WORKSHOP ON WEAR

F. F. Ling

A few facts about ongoing activities at RPI might help to define the scope of the wear problem.

Under a National Science Foundation Project, "Collaborative Research in Tribology" (between Georgia Tech, Northwestern and RPI), RPI is committed to defining the regimes of wear that are important nationally. This is done under the surveillance of an Industrial Advisory Board and the survey is to be in-depth and quantitative; this means going into plants and installations to examine records and worn parts. Concurrently, under the support of the Navy, an ongoing project for several years has led M. B. Peterson to all major Navy rework facilities; the charge there is to examine the nature of predominate tribological failures and recommend solutions. On the scientific front, under the NSF grant, RPI is studying scoring wear as a collaborative program with the other two institutions. At the same time, there are two studies in wear at RPI: (a) Impact wear; (b) Deformation wear. Both studies involve theoretical, computational, and experimental efforts.

As an addendum, a suggestion is made for the future in view of the fact that experimental results on wear are so sensitive to variations in different heats of the same material

and that different conclusions are drawn from individual experimental observations.

. The Committee on Wear of Engineering Material, Organization of European Corporative Development has a single heat of bearing steel from which every researcher is welcome to have a piece. Results from this material will serve as an anchor for comparison of individual efforts, of which there are many around the world. Mr. Donald Buckley of NASA Lewis Research Center is the U.S. representative.

. Among the substantial pieces of work already done, the work at Timken Roller Bearing Company is especially worth noting. What brought them to mind was their work on initiation of cracks including sub-surface ones. Based on metallurgical studies, it has been shown that premium steels which are CV remelted would virtually eliminate sub-surface inclusions which compete with surface flaws as initiators of fatigue. Dr. Walter Littmann should be contacted for these materials.

. Regarding copper, the International Copper Research Association (INCRA) has an ongoing sponsored work at RPI on the wear of copper based material. When released, these data should be useful for those studying the wear of this material system.

EROSION BY SOLID PARTICLES

I. Finnie

In studying the mechanism of erosive wear it is convenient to separate materials into two categories - ductile and brittle. An ideally ductile material undergoes large plastic strains before fracture occurs and volume removal in erosion may occur by the cutting or displacing action of the particle. By contrast, in an ideally brittle material only elastic deformation occurs prior to fracture so volume removal must be due to the propagation and intersection of cracks around the impacting particle. Clearly these concepts may be over simplified but they do allow useful analytical solutions to be developed. That ductile and brittle materials respond to erosion in very different ways can be seen from the variation of erosion with angle of impingement. This is shown schematically in Fig. 1 with the maximum removal being taken as the same. Since the shape of the curve of erosion versus angle is so different any comparison has to be made for a given angle. Also, it turns out that the effects of particle velocity and particle size are different in the two cases.

Taking first the case of ductile behavior one can consider an idealized rigid particle and with certain assumptions write the equations of motion for its path. These are spelled

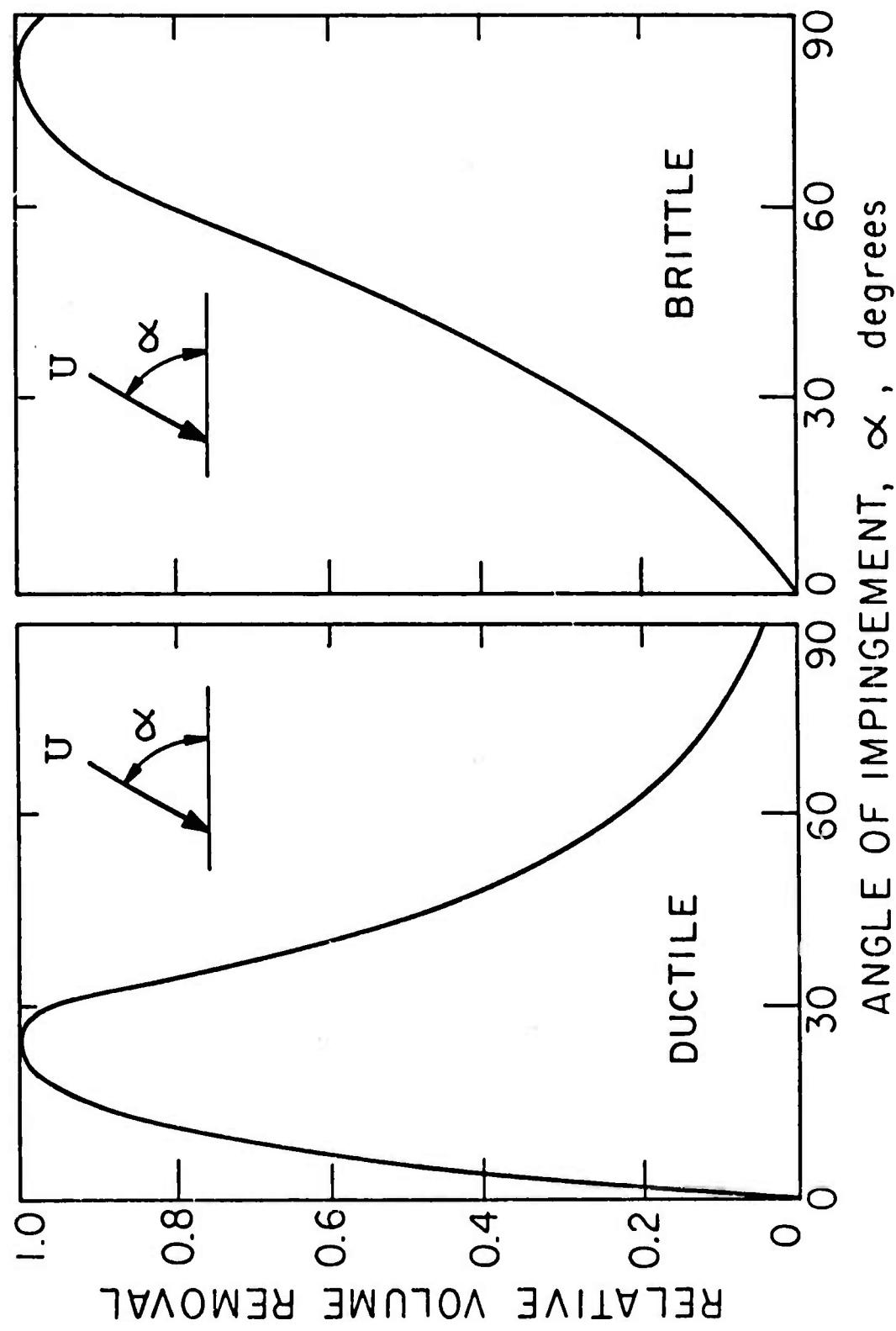


Figure 1. Volume removal as a function of angle shown schematically for ductile and brittle solids.

out in Refs. 1 and 2. The result if we assume the path swept out by the particle tip is the volume removed is:

$$V = \frac{CMU^2}{4p \left(1 + \frac{mr^2}{I} \right)} \left[\cos^2 \alpha - \left(\frac{\dot{x}_t}{U} \right)^2 \right]$$

where

V = volume removed from surface,
 M = mass of eroding particles,
 m = mass of an individual particle,
 I = moment of inertia of particle about its center of gravity,
 r = average particle radius,
 α = angle of impact,
 U = particle velocity,
 p = horizontal component of flow pressure,
 c = fraction of particles cutting in idealized manner,
 \dot{x}_t = horizontal velocity of tip of particle when cutting ceases.

Not all the particles will cut in this idealized manner in fact with abrasive grains only about 10% probably will do this. Some land on flat faces, other indent and pile up material which is later removed or fractures. This type of mechanism indenting rather than cutting may be important at high angles α but I believe that the one I have postulated is the primary mechanism of erosion by abrasive grains at lower angles.

Fortunately erosion is a very inefficient process.

Not only are 90% of the grains loafing, the efficiency which the 10% that cut transfer their energy to the surface is very low. From the equations taking $I \approx 1/3 mr^2$ it can be shown that 3/4 of the initial kinetic energy of translation goes into kinetic energy of rotation with only the remaining 1/4 doing plastic deformation. The analysis predicts $V \sim U^2$ and early tests appeared to confirm this result. However, more careful velocity measurements have shown exponents greater than 2. Typical values range from 2.2 to 2.4 with values as high as 3 having been reported.

We now examine the predicted variation of erosion with angle. One condition for the end of cutting is the horizontal velocity of the particle's tip becomes zero. This leads to $V \sim \cos^2\alpha$. However at low α the particle may leave the surface while still moving horizontally. This leads to the result $\dot{x}_t'/U = \cos\alpha - \frac{2}{P} \sin\alpha$ where $P = K : (1+mr^2/I)$. Maximum erosion is predicted to occur when $\tan 2\alpha = P$. Experimental results shown in Fig. 2 show good agreement with prediction for low angles. The idealized cutting analysis isn't good for large α and one can postulate other removal mechanisms in this range.

An intriguing aspect of erosion of ductile metals at low angles is that a series of ripples forms on the surface. One can explain this quantitatively by repeating the analysis for surface with curvature³. It is seen that when the end of cutting is determined by the particle leaving the surface,

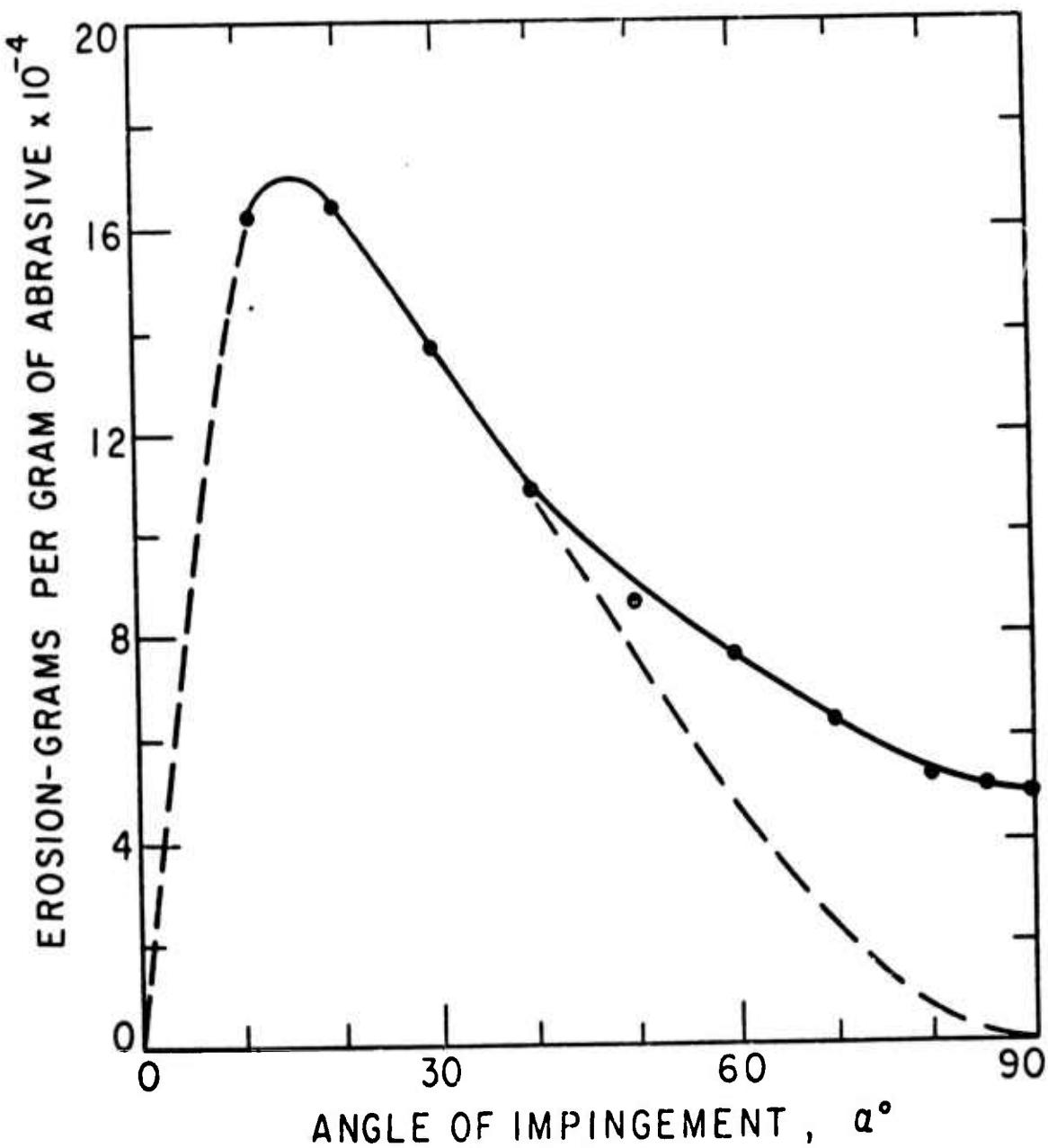


Figure 2. Experimental data for 1100-0 Al eroded by 120 mesh (0.004 inch) SiC particles at 500 ft/sec. Predicted (dashed) variation with angle is scaled to give the same maximum erosion in both cases.

peaks should erode less than valleys. Thus we expect any surface irregularities to accentuate when erosion occurs at grazing angles.

The analysis predicts no effect of particle size. However, it is found experimentally, that as particle size decreases below say 70 to 100μ there is a rapid decrease in the rate of erosion. Fig. 3 shows a typical result. The result is fortunate in that it is hard to get rid of small particles in separators but its explanation is baffling.

The similarity of the two curves in Fig. 3 indicates that the mechanism of erosion is the same for small and large particles. One explanation advanced is that the larger particles fragment and produce secondary damage. However the same size-effect is found in slow speed abrasion tests. We have found this same size-effect in eroding single crystals, in eroding gold (no oxide) and can only conclude that there is a true physical size effect such that small regions show higher hardness values. This is consistent with hardness tests made under a scanning microscope which showed about a threefold increase in hardness relative to bulk specimens⁴.

What role does the surface play. The simple approach would be to say that for annealed metals the flow pressure p would be proportional to Vickers hardness. Thus $V \sim (Hv)^{-1}$. Figure 4 shows that for several annealed fcc metals (and Ta) this relation is followed very closely. Departures are shown by b.c.c. materials. This is not surprising since Vickers

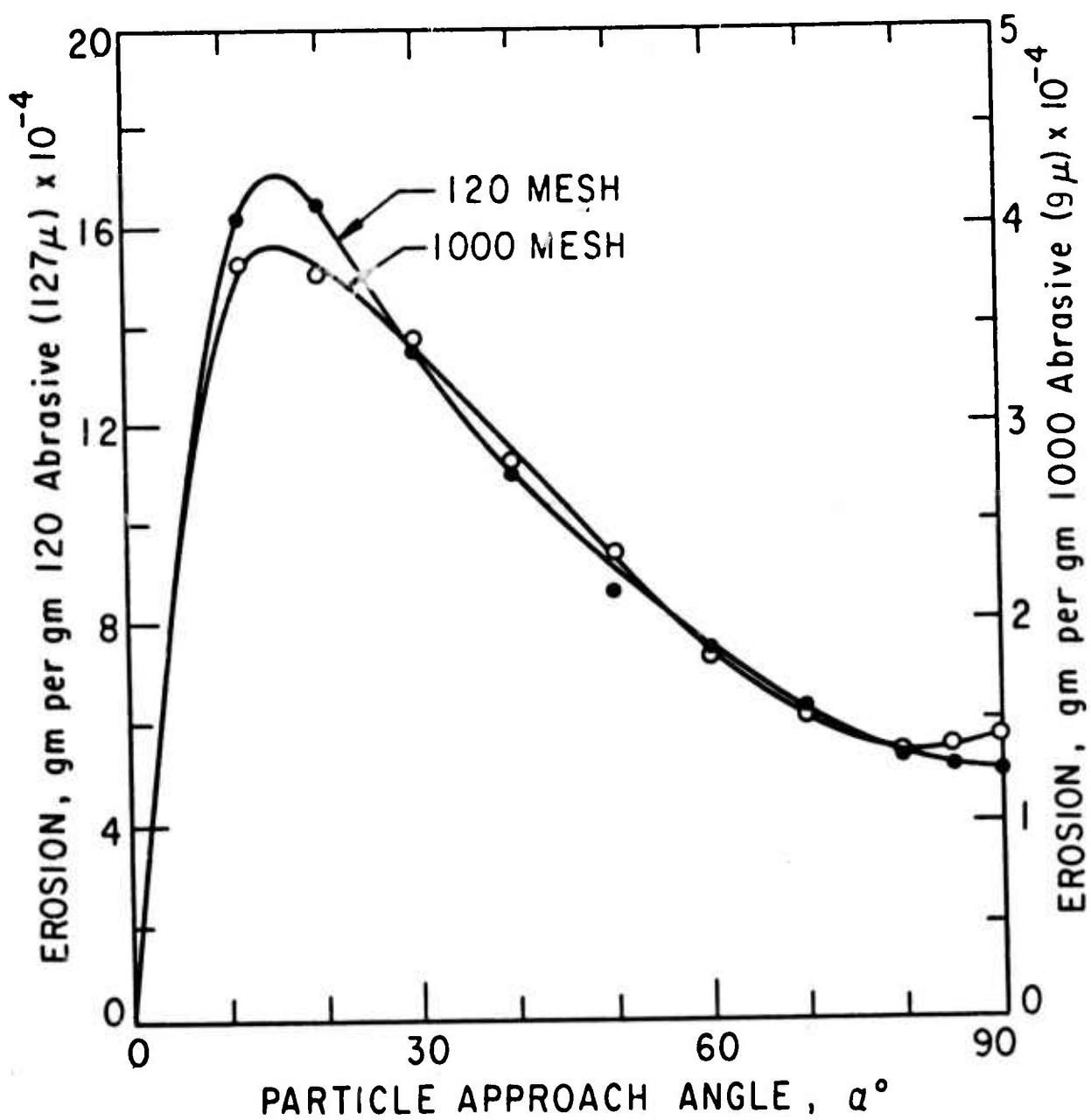


Figure 3. Weight removed when 1100-0 Al is eroded by SiC particles of two sizes at 500 ft./sec.

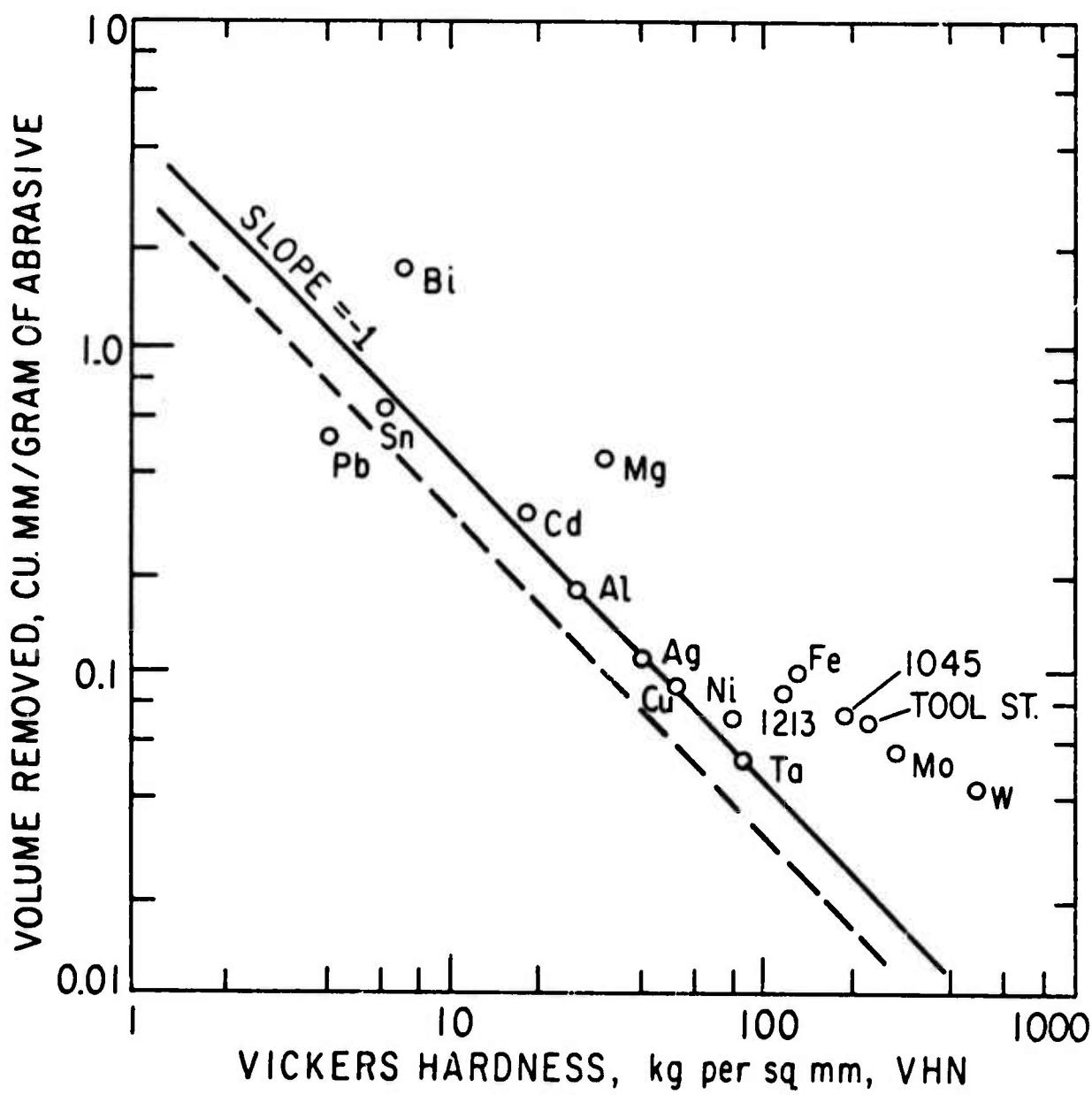


Figure 4. Volume removed as a function of Vickers hardness when annealed metals are eroded by 60 mesh SiC at $\alpha = 20^\circ$ and $U = 250$ ft/sec.

hardness is a measure of plastic flow at low strains (approx. 8%) and erosion will involve high strains. Unless all these metals have geometrically similar stress-strain curves, we wouldn't expect an exact relation between Vickers hardness and flow pressure in erosion. The dashed line is what we predict from the analysis for $c = 0.1$. A useful replot in Fig. 5 is to show the inverse of the volume removed which might be termed the erosion resistance. Here we see the discouraging result that changing hardness by cold work or by heat treatment has little effect on erosion resistance. This confirms the practical experience that when pure erosion under oblique impact is involved, ordinary steel is as good as anything else. By contrast in abrasive wear, although cold work has little effect on wear resistance, heat treatment is beneficial.

Erosion of Brittle Solids

When we turned some years ago to examine the erosion of brittle solids it was not as easy to get started as with ductile metals. One thinks first perhaps of an energy approach but the essential energy requirement - that to produce fresh surface is, in all crushing operations, a small fraction of the energy required. Studies of chipping under a sharp wedge have been made by people in rock mechanics and coal mining - the equations they derive are exactly those suggested for metal cutting 80 years ago - but do not seem appropriate for erosion by polyhedral abrasive particles.

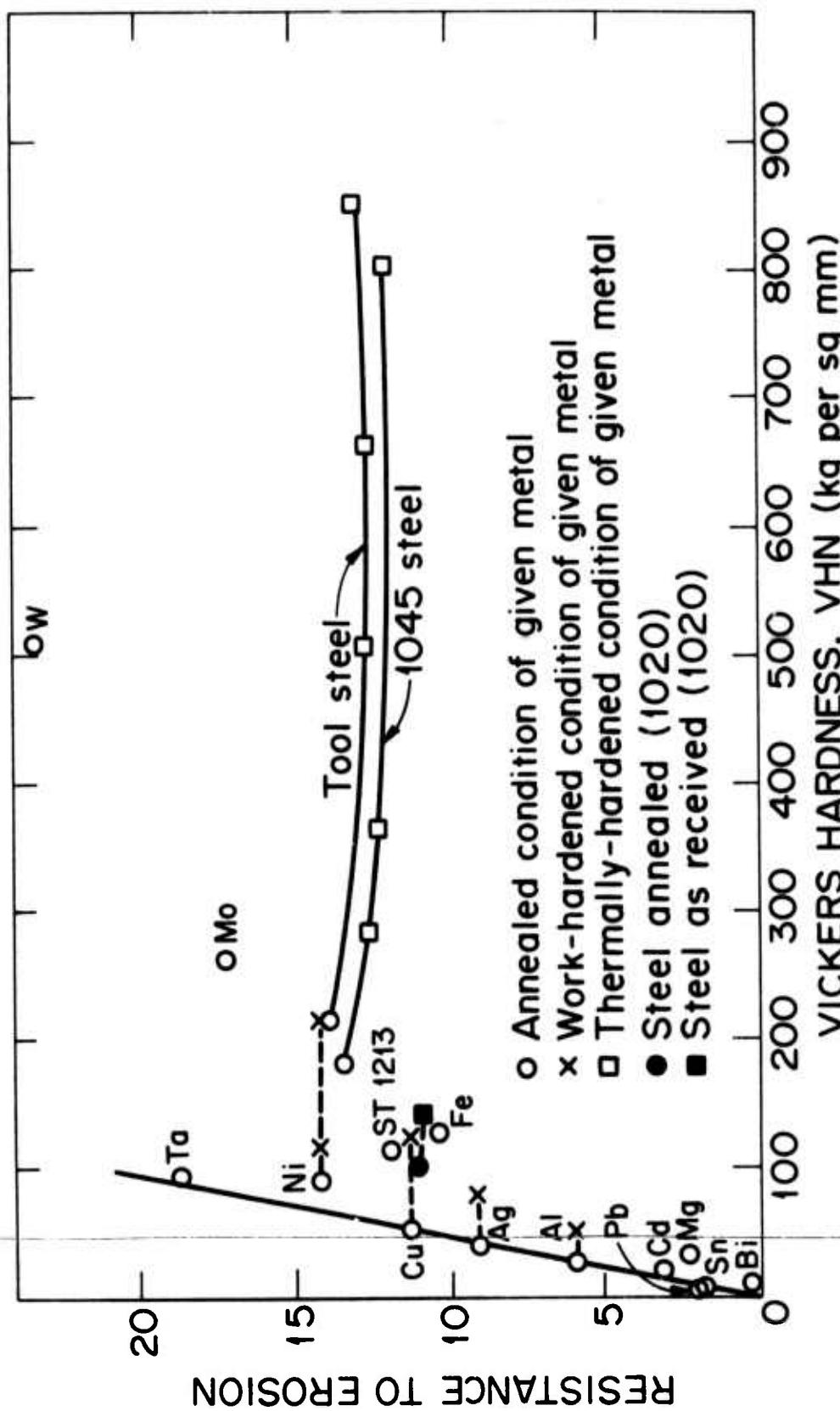


Figure 5. Resistance to erosion (grams of eroding particles removed) as a function of VHN for erosion by 60-mesh silicon carbide particles at $\alpha = 20$ deg and a velocity of 250 ft/sec.

When we examined a number of brittle solids eroded by spheres or grains the results could be written as $V \sim U^a d^b$ ($\alpha=90^\circ$). At first there seemed no relation between these parameters and such properties as modulus and mean strength. However, we then recalled the Weibull statistical treatment of brittle strength which predicts a strong size effect. The higher strength values which are observed when smaller regions are tested arise because of the decreased probability of finding severe flaws in small regions. Thus, an increase in velocity or particle size both increase the loaded region and should result in lower strength values. By turning to this Weibull treatment we were able to explain a number of features of brittle erosion. We consider first spherical particles which are easier to study. More predictions are possible than for angular abrasive particles where we have to be considerably more empirical.

Weibull's approach, in essence, states that the probability of failure may be written as

$$F(\sigma) = 1 - e^{- \int \left(\frac{\sigma - \sigma_u}{\sigma_0} \right)^m \frac{dV}{V_0}} ; \quad \sigma > \sigma_u$$

$$= 0 \quad \quad \quad \sigma \leq \sigma_u$$

The integral is taken over the region stressed in tension. Normally, the volume under tensile stress is considered but for glass in which the strength impairing flaws are confined to the surface an integration over the area stressed in tension

is more appropriate. V_0 denotes the unit volume (or area) for which the parameters σ_u , m and σ_0 are obtained. The parameter σ_u ("zero strength") is sometimes taken as zero to give the simpler two parameter Weibull distribution.

Taking Weibull's ideas we have shown that by estimating m , σ_u and σ_0 from bending tests we can predict many aspects of the fracture that occurs under a spherical indenter⁵. The actual fracture pattern looks as shown in Fig. 6. The important point is that we can estimate the extent of cracking due to load P or velocity U for a single particle. Many concentric ring cracks form and by extending Weibull's treatment to predict fracture location we can predict the average diameter of the first and outermost cracks. The initial cracking is statistical in nature but the depth of the crack is a problem in fracture mechanics which one can predict in principle knowing the fracture toughness of the material. However to a first approximation the depth S of cracking in a given material is proportional to the indentation that one calculates from the Hertz equation. Anyway, by assuming that the region damaged by a single particle is say proportional to the damage produced by one of many particles we can make a number of predictions. The effect of compressive stress is essentially to change the parameter σ_u and this can be studied experimentally (Fig. 7). The volume removed per particle can be measured and compared with predictions as in Fig. 8. Here $rU^{4/5}$ is proportional to the depth of indentation. One could perhaps be more quantitative

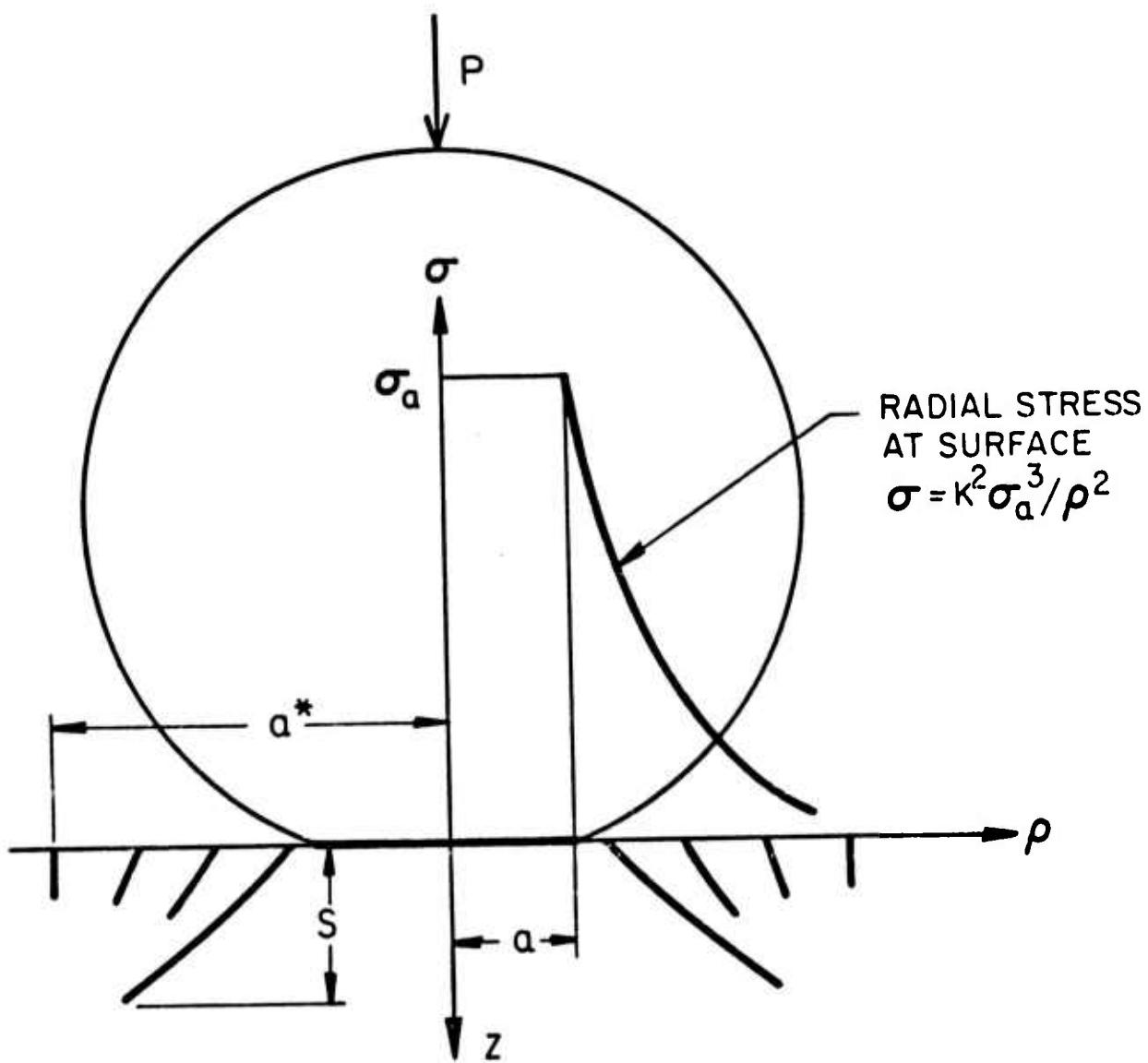


Figure 6. Coordinate system used in study of ring cracking.

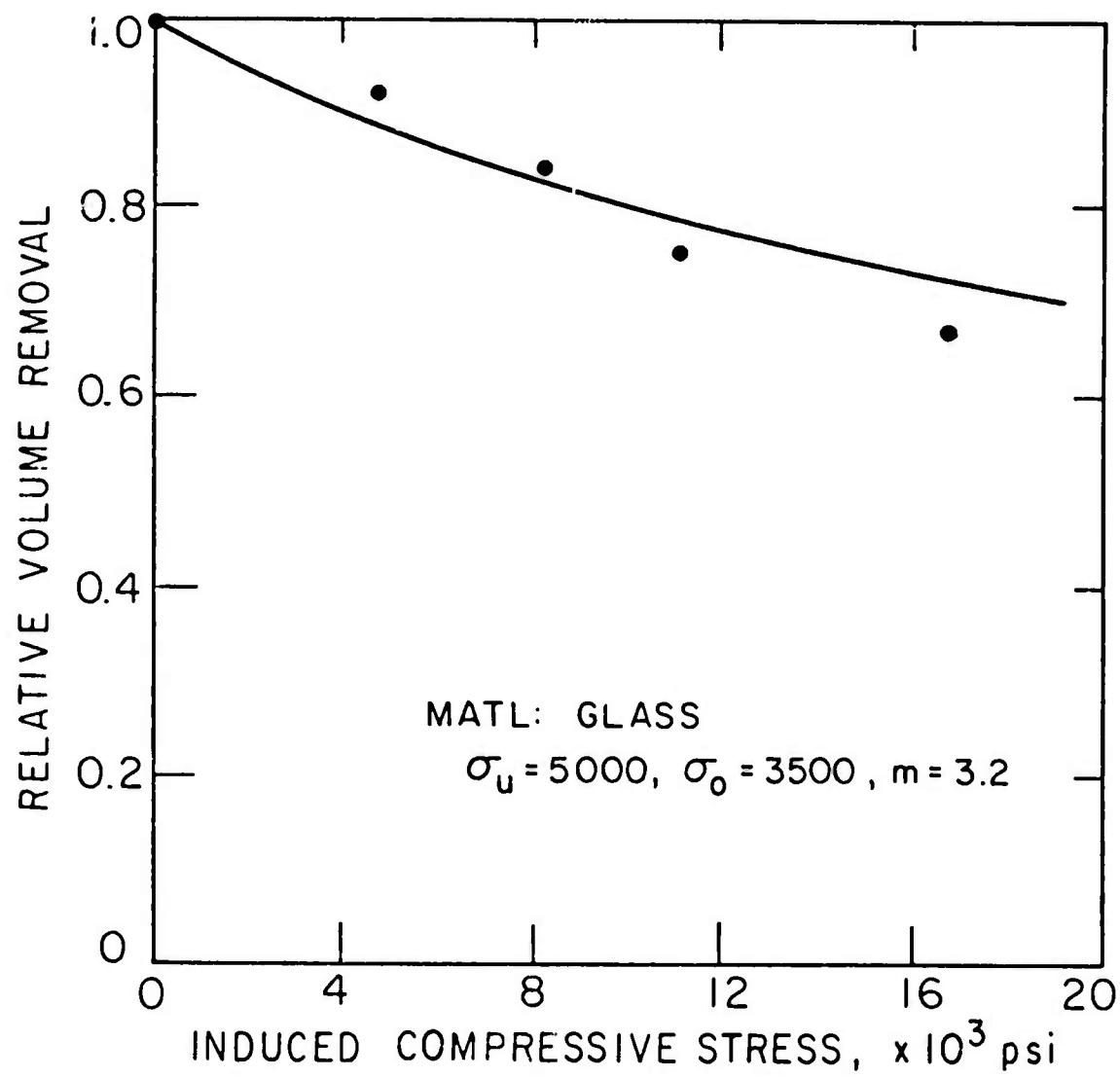


Figure 7. Influence of biaxial compressive stress on erosion of glass by 0.0165 in diameter steel shot at $U = 270$ ft/sec. and $\alpha = 90^\circ$.

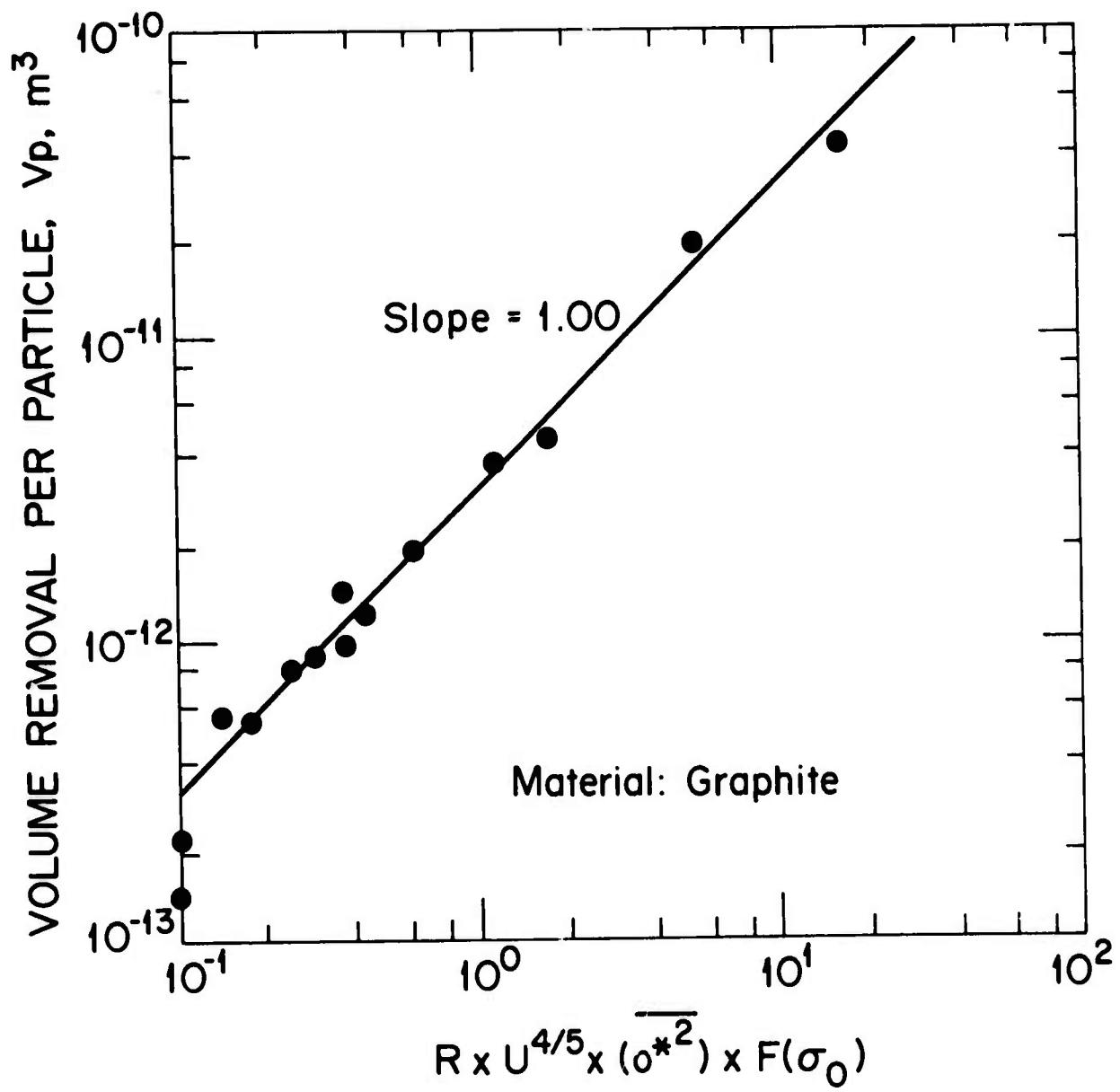


Figure 8. Experimentally observed values of volume removed per particle (for a wide range of velocity and particle size) as a function of the predicted damaged volume. Material: graphite.

tive and try to get the constant of proportionality from studies of the depth of the ring crack and in fact it is of the order of unity. The analysis can be extended to erosion by oblique impact and predicts the shape of the curve quite well as seen in Fig. 9.⁶

For angular abrasive particles we have had to take another approach. We assume as shown in Fig. 10 that the cutting is done by a small indenter of size r' which is independent of the particle size r . A similar expression is written for the size of the damaged region but we take the radius r^* as that of a spherical particle which just produces ring cracking when indenting to the depth x . x is calculated for a sphere of r' with mass given by r . The analysis is made for a two parameter Weibull distribution and the details are given in Ref. 6. The results shown in Fig. 11 are quite encouraging with only the case of glass eroded by steel shot showing serious deviation from the predictions.

The Brittle-Ductile Transition

An interesting consequence of the size-effect on strength of brittle solids is that if they are loaded on smaller and smaller regions, the stress required to produce fracture may eventually exceed that required for yield. Thus, we can understand the familiar observation that microhardness or scratch hardness tests on brittle solids may produce flow rather than fracture. In erosion the response of brittle and ductile solids to changes in the angle of impingement are

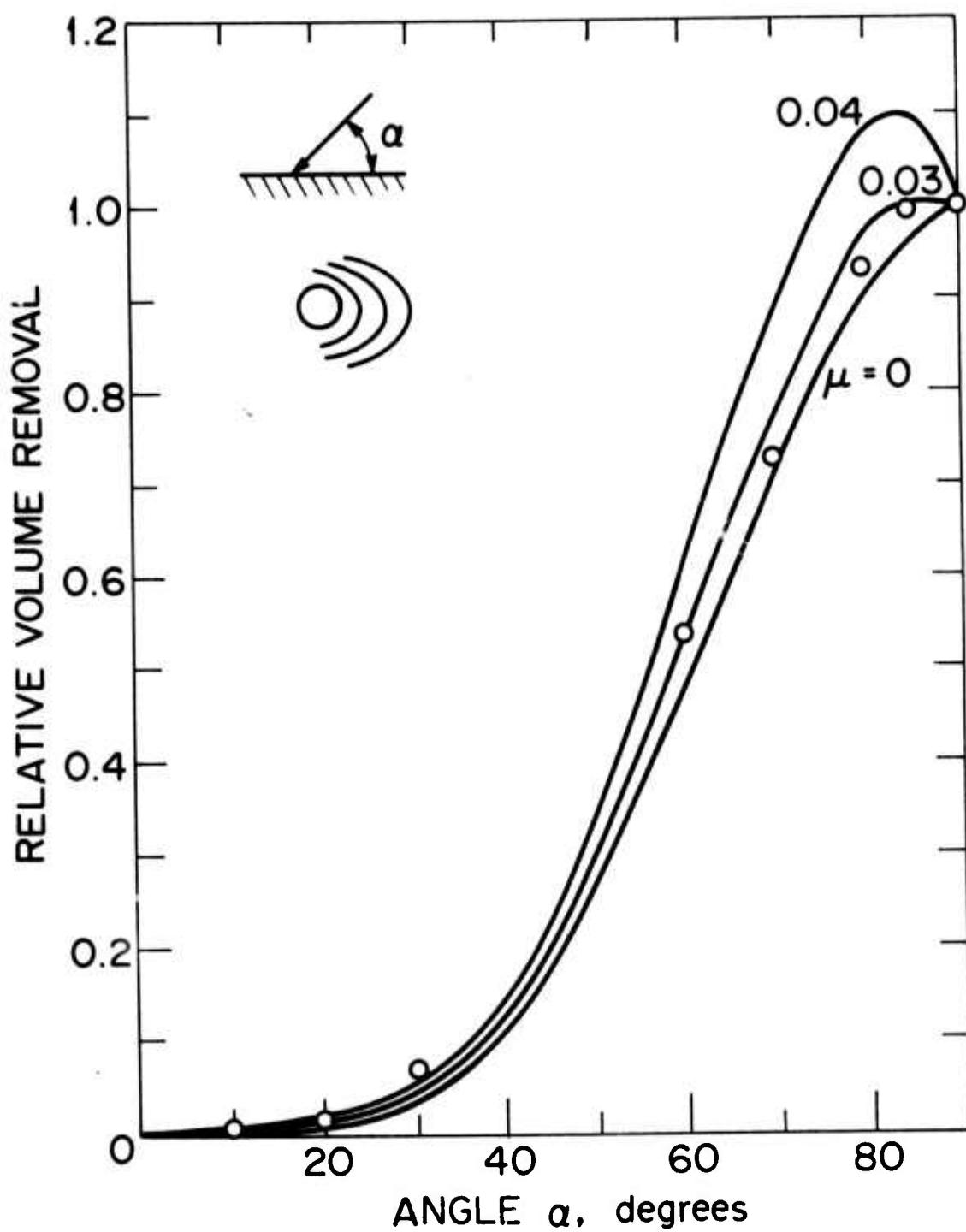


Figure 9. Predicted and observed values of the relative volume removal as a function of angle of impingement for glass eroded by No. 110 steel shot at 200 ft/s (60.96 m/s). Predictions are made for three values of μ , the coefficient of friction. The inserted sketch shows the contact circle and the horseshoe cracks.

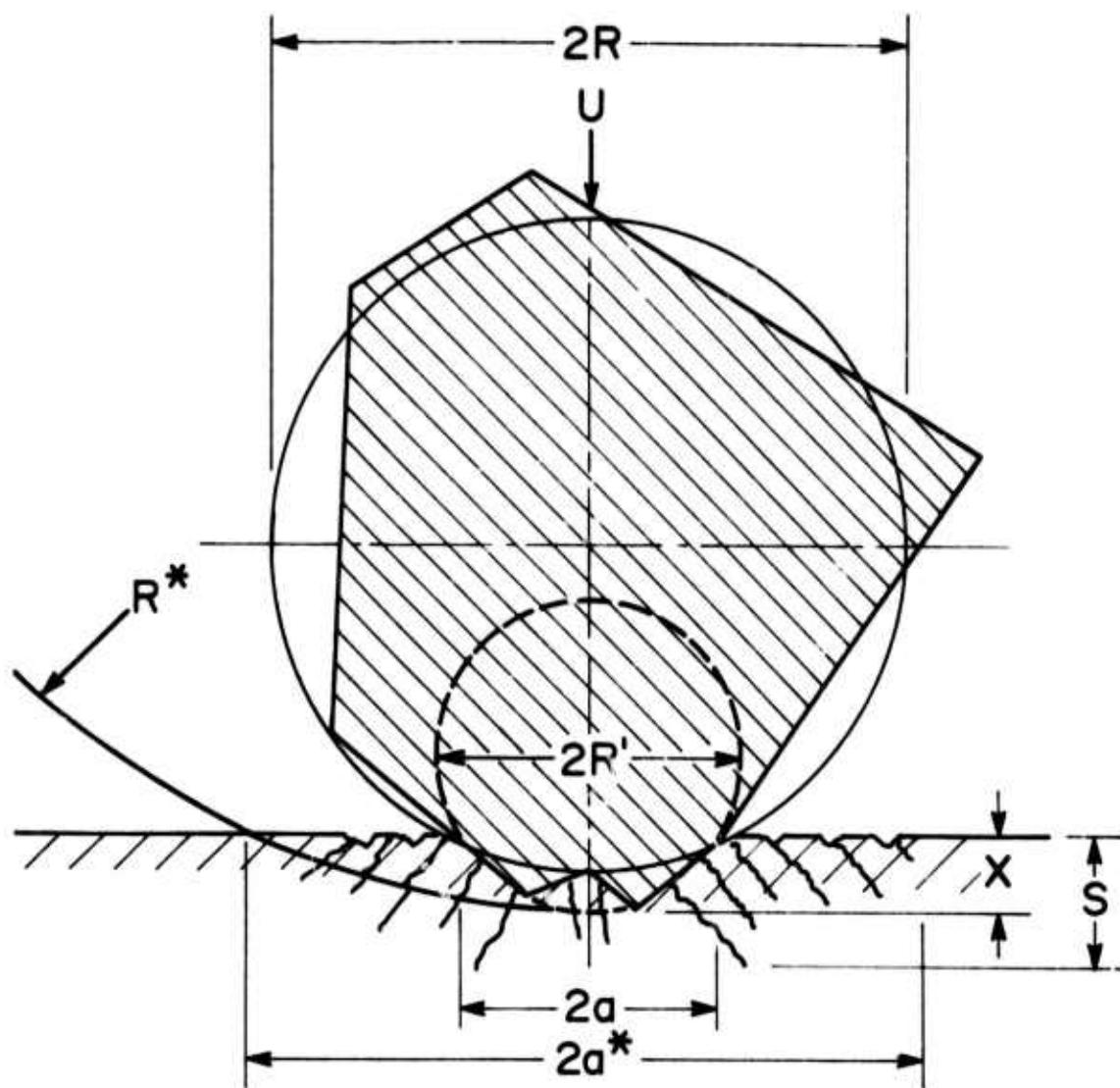


Figure 10. Schematic view of cracking due to particle impact.

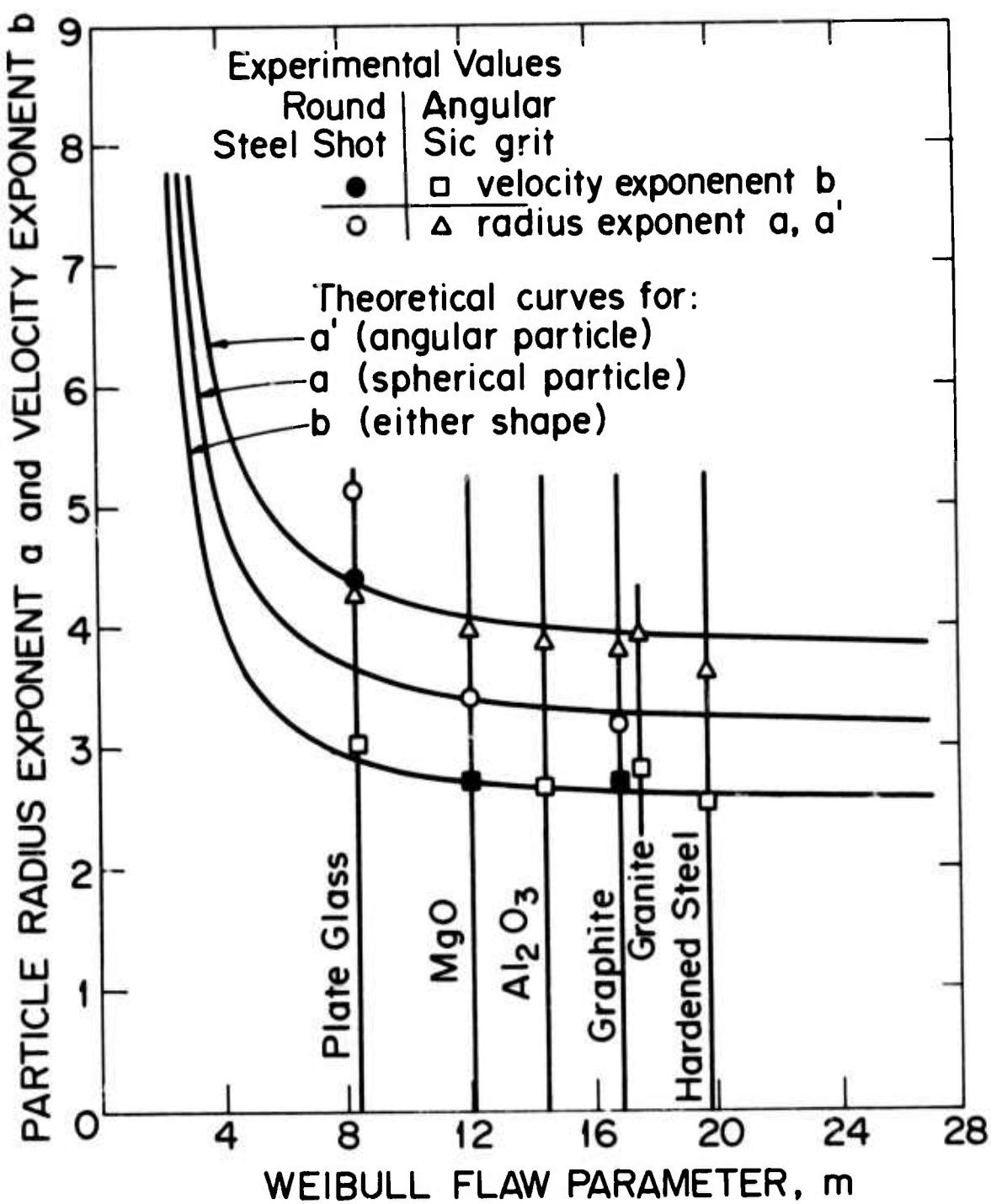


Figure 11. Comparison of theoretical and experimental values of particle velocity and radius exponents as a function of the Weibull flaw parameter m .

markedly different. Thus a transition to ductile behavior should be easily detectable.

To estimate the particle size at which ductile behavior should be noticed in erosion testing of nominally brittle solids, we need only the parameters of the Weibull distribution and indentation hardness measurements.

Again the details are given in Ref. 6 and predict that a transition to "ductile" behavior should be observed for particle sizes somewhere between 2 and 20 microns. Figure 12 shows that such a transition is observed in erosion tests. In addition, ripples are observed on the glass surface eroded by 9 micron particles which is an additional indication of ductile behavior.

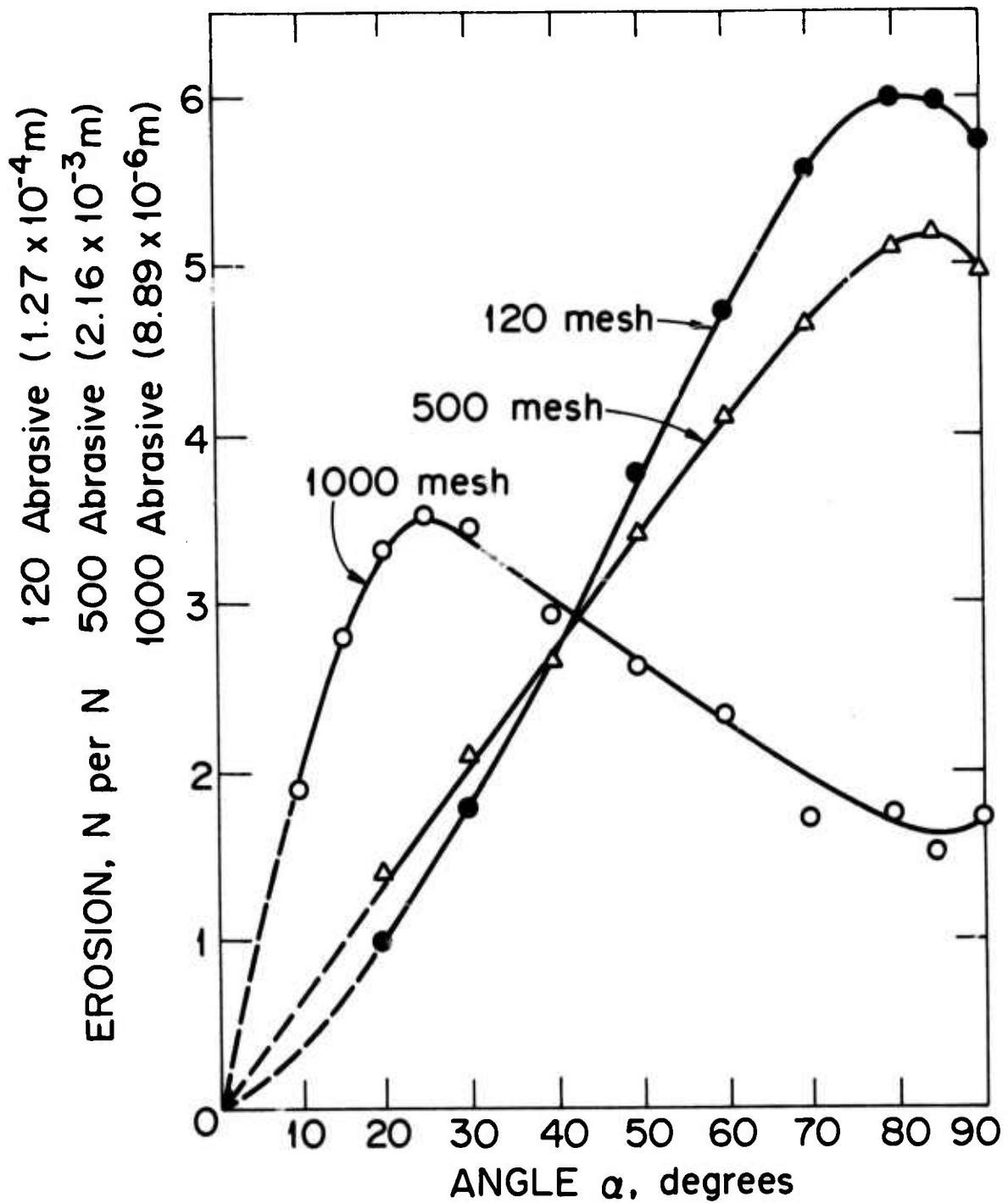


Figure 12. Weight removal per newton of abrasive as a function of particle approach angle α for glass eroded by silicon carbide particles at 500 ft/sec (152.4 m/s).

References

1. Finnie, I., The Mechanism of Erosion of Ductile Metals, Proc. 3rd Nat. Congress Appl. Mech. 527-532 (1958).
2. Finnie, I., Erosion by Solid Particles in a Fluid Stream, "Symposium on Erosion and Cavitation," STP 307 ASTM 1961 pp. 70-82.
3. Finnie, I. and Kabil, Y. H., On the Formation of Surface Ripples during Erosion, Wear 8, 60-69 (1965).
4. Gane, N. and Cox, J. M., The Micro-hardness of Metals at Very Low Loads, Phil. Mag. 22, 881-891 (1970).
5. Oh, H. L. and Finnie, I., On the Location of Fracture in Brittle Solids - I, Due to Static Loading, Int. J. Fract. Mechanics 6, 287-300 (1970).
6. Oh, H. L., et al, On the Shaping of Brittle Solids by Erosion and Ultrasonic Cutting. "The Science of Ceramic Machining and Surface Finishing," U.S. Nat. Bureau of Standards Spec. Pub. 348, pp. 119-132.

THE THERMODYNAMIC CRITERION FOR CRACK GROWTH

R. M. Thomson

ABSTRACT

Recent experimental results show a generally logarithmic dependence between the pressure of atmospheric hydrogen and the stress intensity factor, K . We show that the experimental results can be rationalized in terms of the decrease of surface energy by adsorbed hydrogen. Theoretical expressions for the effect on surface energy are given. Quantitative comparison, however, must await working out the functional relationship between the true surface energy and the effective surface energy which includes effects of the plastic zone surrounding the crack.

THE THERMODYNAMIC CRITERION FOR CRACK GROWTH

R. M. Thomson

INTRODUCTION

In this paper we present a thermodynamic theory for the threshold (often called the static fatigue limit) for slow crack growth in a relatively brittle material which is in the presence of an embrittling gaseous environment. (In the present paper, only a part of the argument will be presented. The complete theory will appear in a paper jointly sponsored by the NBS.) We shall apply it primarily to the embrittlement of steel by gaseous external hydrogen, although the theory is in no way limited to this particular choice.

Our approach looks first to the lowering of the surface energy of the faces of the crack caused by the presence of hydrogen, and then to the functional connection between the change in surface energy and the change in the size and character of the plastic zone surrounding the crack. In spirit, we thus return to the original classic philosophy of Griffith¹ in 1921 for the prediction of crack growth threshold. Indeed, the onset of fracture must be governed in the last analysis by thermodynamics, and one is of course not required in the thermodynamic case to worry out the enormously difficult problem of kinetics for the various complex processes

which can occur at the crack tip and in its surrounding plastic zone.

Practically, of course, the slow crack growth regime may be governed by processes which approach zero velocity very slowly, so that the practical threshold at which the crack velocity has reached some preselected low value is not the theoretical threshold in our sense. Our thermodynamic threshold thus becomes a lower bound for the various ways of defining the fracture threshold.

The thread of our argument involves two major lines of discussion in modern fracture theory. The first relates to the fundamental character of the fracture process itself, and the second concerns the role of hydrogen in the phenomenon. We discuss these points in turn before turning to the analysis.

MODELS OF FRACTURE

Fracture can occur by either of two generally separable mechanisms. In the first, new surface is formed and the crack advances primarily by plastic processes, and hole growth is the most common manifestation. In the second form, which we shall term classic fracture, and which is also called lattice decohesion, fracture occurs by the continuous opening of a crack which is sharp on an atomic scale, but which is generally surrounded and shielded by a macroscopic plastic zone. Mixtures of these two poles of behavior can and do exist. For example, the initiation of a hole in the intense plastic zone

in front of a crack may be by means of classic fracture events at precipitate or grain boundaries. We suggest that hydrogen embrittlement, in its more brittle regime is a case to which the classical fracture model can be applied.

We thus view plastic fracture and classical fracture as competing mechanisms, but there is one condition under which only plastic fracture can occur. When the shear stress at the tip of a crack is sufficiently high to generate dislocations spontaneously, then any atomically sharp crack will decompose and classic fracture is impossible.² Rice and Thomson estimate that iron is a borderline case for spontaneous dislocation formation, and the computer simulation calculations in iron of Kanninen and coworkers³ also suggest that something akin to dislocation formation occurs at a crack tip at the critical stress. However, since the surface energy is lowered when the iron fracture surface is covered with hydrogen, the threshold stress for crack growth is also lowered, and classical fracture again becomes possible.

In our model, we envision the atomically sharp crack to coexist with its surrounding plastic zone because the dislocation density is bounded by the strain hardening and dynamic recovery properties of the crystal, and the dislocation distribution is highly inhomogeneous because the dislocations form into a gross cellular structure. Within the interior of a given cell, the crystal is relatively dislocation free with a high degree of local order over distances of the order of

microns. Hence, in this case, blunting reactions by dislocations at a crack tip are relatively rare events on an atomic scale, and one would expect a classical fracture tip to be stable against the kind of smoothly rounded plastic blunting envisioned in a plastic continuum theory. Such a sharp crack will interact strongly with the dislocation cell walls, and with any stray dislocations it meets in a blunting configuration, but we postulate that the interaction does not destroy its basic character as a sharp crack. In this connection, we expect the crack as it moves through the crystal will flow around and over obstacles in much the same way as dislocations do in the flexible string model.

Our model thus poses a question which is quite subtle, namely what are the quantitative conditions under which dislocations can be generated in sufficient numbers and on such a fine distribution in the plastic zone as to round out and destroy a classical fracture. In addition to the case discussed by Rice and Thomson where spontaneous generation occurs, cases probably also exist where the plastic zone also exhibits sufficient plastic instability on a sufficiently fine scale as to blunt out any classic fracture configuration down to atomic dimensions.

HYDROGEN EMBRITTLEMENT

Investigators of hydrogen embrittlement have for many years fallen into two general camps, those who believe that

the embrittlement process is one which occurs at the crack tip itself, and those who believe that the embrittlement process is due to a transport into critical regions deep within the crystal.

The most recent and articulate spokesman for the former crack tip view is Oriani and coworkers³. In essence, Oriani maintains that the Orowan stress given by the condition that the bonds in the cohesive region near the tip are stretched to their maximum force is modified by the presence of the hydrogen in a uniform way. This condition is the same as that for dynamic fracture, and the slower velocity in the case where hydrogen is present is presumably determined by the diffusion of the hydrogen out of the gaseous phase into the cohesive region of the crack. This model is thus not a thermodynamic model, but rather a kinetic model. We believe it should be possible for a bond to break at the tip by thermal fluctuations even though it is not stretched to the Orowan limit, if the Griffith thermodynamic condition is satisfied. Nevertheless, our model is very closely related to that of Oriani in that we shall assume that the crucial role of the hydrogen is played at the newly opening surface at the crack tip rather than in the interior.

The second view, that the hydrogen must penetrate deeply into the plastic zone before embrittlement can occur is essentially a proposition that hole growth is enhanced by hydrogen. This view of hydrogen embrittlement is the one most

generally held in the metallurgical community, and the fact that the maximum lattice dilatation around a blunted crack occurs at some distance inside the plastic zone is often used to estimate the location of the critical region for hole initiation. In this work, we will assume that the intergranular fracture seen in the higher strength steels corresponds to a brittle fracture mode to which our model is applicable.

SURFACE ENERGY EFFECTS

In this section we calculate the change of surface energy caused by the external hydrogen gas. Although Petch has given some equations for this effect earlier,⁵ we shall rederive these equations in a more careful manner in order to discuss the difference between H_2 and D_2 as measured by Oriani and Josephic³.

The thermodynamic criterion for the fracture threshold is given by the equation,

$$\delta F_{\text{total}} = 0 = \delta F_{\text{elastic/plastic}} + \delta F_s + \delta F_g$$

F_{total} is the free energy of the total system, $\delta F_{\text{elastic/plastic}}$ is that for the elastic and plastic portions of the system, including any hanging weights, δF_s is the surface contribution and δF_g is the free energy change in the external gas. If the crack advances by one atomic spacing all along its front, we shall write $\delta F_{\text{elastic/plastic}}$ as G , the total crack extension force, which we have analyzed in a subsequent work, and will

not further discuss here. (The G we have defined, of course, is not the normal definition and differs from it by plastic terms.) When the crack moves forward by one atomic spacing, the surface and gaseous free energies change by the amounts,

$$\delta F_g = -\frac{1}{2}\mu_g \Gamma (2a_0)$$

$$\delta F_s = 2\gamma a_0 + 2a_0 \Gamma \mu_s$$

It is assumed that the crack is a line one cm in length, and advances one atomic spacing all along its length. a_0 is the lattice spacing. Γ is the number of H atoms on the surface per cm^2 , assumed to be chemisorbed as atomic hydrogen. μ_g is the chemical potential of molecular hydrogen in the gas, and μ_s is the chemical potential of hydrogen atoms on the surface. In equilibrium $\mu_s = \frac{1}{2}\mu_g$. Thus at threshold,

$$G = \gamma$$

where γ is the true thermodynamic surface tension of the covered surface.

When hydrogen is present, the Gibbs adsorption equation is

$$\begin{aligned} d\gamma &= -S_s dT - \Gamma d\mu_s \\ &= -S_s dT - \Gamma \left. \frac{\partial \mu}{\partial T} \right|_P dT - \Gamma \left. \frac{\partial \mu}{\partial P} \right|_T dP \\ \gamma &= \gamma_0 - \frac{1}{2} \left. \frac{\partial \mu_g}{\partial P} \right|_T \Gamma dP \end{aligned}$$

γ_0 is the reference γ for the uncovered surface at temperature

T. We shall assume that γ_0 does not vary with temperature.

For the gaseous hydrogen,

$$\frac{\partial \mu_g}{\partial P} \Big|_T = \frac{kT}{P}$$
$$\gamma_0 = \gamma_0 - \frac{kT}{2} \int_0^P \frac{P(T)}{P} dP$$

To find $\Gamma(T)$, we shall adopt the Langmuir equation, although where an empirical isotherm is available, it could also be used. To obtain the Langmuir isotherm, we note that the partition sum of the adsorbed gas on the surface when it is immobile on that surface is

$$Z_s = Z_{\text{conf}} Z_g$$

Z_g is the partition sum of the single atom on the surface relative to that atom bound in a gaseous molecule, and Z_{conf} is the configurational part of Z . We find that

$$\ln Z = N_0 \ln N_0 - (N_0 - n) \ln (N_0 - n) - n \ln n + nE_b/kT$$

where N_0 is the total number of surface sites, and n is the number of covered surface sites. E_b is the binding energy of a hydrogen atom to the surface referred to its state in a free molecule. Since the free energy of the surface per unit of surface, f_s , is given by

$$f_s = -kT \ln Z_s$$

and the chemical potential of H atoms on the surface is

$$\mu_s = \frac{\partial f_s}{\partial n}$$

and $\mu_g = 2\mu_s$, we finally have

$$\mu_g - 2E_b = kT \ln \left(\frac{\theta^2}{(1-\theta)^2} \right), \quad \theta = \Gamma a_0^2$$

We also know μ_g for a perfect gas of diatomic molecules. It is given by

$$\mu_g = -kT \ln \left\{ \left(\frac{2\pi mkT}{h^2} \right)^{3/2} \left(\frac{kT}{P} \right) \zeta_{\text{Rot}} \right\}$$

m is the mass of the molecule, h is Planck's constant, ζ_{Rot} is the partition sum for the rotator. For H_2 and D_2 , the rotator partition sums are different, and we have, at $T=298^\circ\text{K}$,

$$\zeta_{\text{Rot}} = \begin{cases} 3.04 & H_2 \\ 4.68 & D_2 \end{cases}$$

The Langmuir isotherm then becomes

$$\frac{\theta^2}{(1-\theta)^2} = bP, \quad b = \left(\frac{2\pi\hbar^2}{mkT} \right)^{3/2} \frac{1}{kT} \frac{e^{2E_b/kT}}{\zeta_{\text{Rot}}}$$

and the surface tension becomes

$$\gamma = \gamma_0 - \frac{kT}{a_0^2} \ln(1 + \sqrt{bP})$$

The important experimental comparison for us is the series of measurements of the hydrogen threshold as a function of pressure by Oriani and Josephic. Measurements were also made of embrittlement by D_2 . Although our theory is not complete without the incorporation of the plastic terms, these

mainly have the effect of multiplying γ by a large constant. In this truncated form of the theory, we can therefore test a portion of the model by checking the form of the pressure dependence, and the effect of D_2 . In Figure 1, we plot Oriani and Josephic's results. Since the crack extension force is proportional to the stress intensity factor, K , we display the results in a semilog plot versus K^2 . In general the results show the expected form except for high values of K , where Oriani and Josephic note that the mode of fracture changes. We also note that the D_2 results are displaced above the corresponding values of H_2 by about a factor of 3. The theoretical factor is 4.35. The theoretical description is thus in accordance with the experimental values.

ACKNOWLEDGEMENT

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHCl5-71-C-0253 with The University of Michigan.

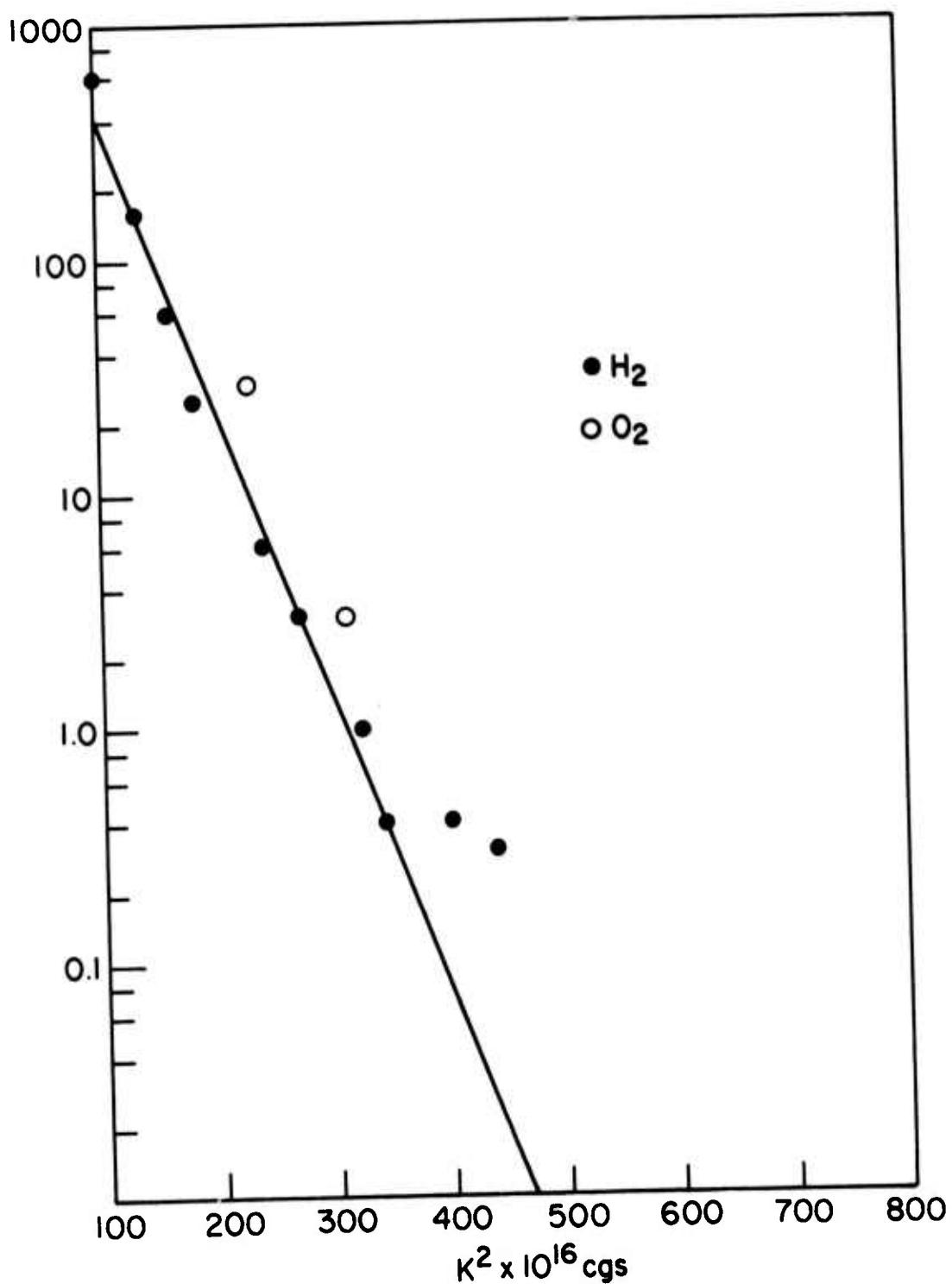


Figure 1. Experimental results of Oriani and Josephic for threshold of crack growth as a function of pressure H₂ and O₂. Experiments performed at room temperature on 4340 steel.

References

1. A. Griffith, Phil. Trans. A221, 163 (1920).
2. J. Rice and R. Thomson, Phil. Mag. 29, 73 (1974).
3. M. Kanninen and M. Gehlen, Int. J. Fract. Mech. 7, 471 (1971).
4. R. Oriani and P. Josephic, Acta Met. 22, 1065 (1974).
5. N. Petch. Phil. Mag. 1, 331 (1955).

NOTES ON THE MECHANICS AND THERMODYNAMICS
OF BRITTLE INTERFACIAL FAILURE IN THE
PRESENCE OF A MOBILE SPECIES (HYDROGEN)

J. R. Rice

ABSTRACT

The thermodynamics of interface separation in presence of a mobile species (H) is discussed. It is argued that low pressure hydrogen in steels involves the weakening of internal interfaces. In some cases these can be grain boundaries and the mode of fracture is then predominantly intergranular; in other cases the critical interfaces are those of inclusions or precipitates, and the role of H is then simply to allow the early nucleation of voids, which spread by conventional transgranular ductile rupture. Attention in the thermodynamic discussion is directed to the limiting cases of separation at constant potential of the solute, and of separation at constant surface concentration. Effects on both the separation energy and the maximum cohesive strength are discussed.

NOTES ON THE MECHANICS AND THERMODYNAMICS
OF BRITTLE INTERFACIAL FAILURE IN THE
PRESENCE OF A MOBILE SPECIES (HYDROGEN)

J. R. Rice

INTRODUCTION

The fracture resistance of ductile alloys is largely determined, in materials or circumstances that do not allow crystalline cleavage, by the strength of internal interfaces. For example, in the presence of temper-embrittled conditions, impurities segregated on grain boundaries allow a brittle decohesion there, and intergranular fracture results. Alternatively, whether assisted by segregated impurities or not, the more normal situation is that interfaces between inclusions or precipitates decohere, and the resulting voids grow to coalescence by the normal plastic flow processes of ductile rupture.

It is striking that hydrogen induced embrittlement of steels, at least under low equilibrating potentials for H, does not seem to involve any new mode of fracture. Interfaces again seem to play the dominant role, and it can be argued that the presence of H has simply made easier fracture mechanisms of a kind which occur normally in such materials.

Thus, in developing a theory of HE (hydrogen embrittlement), it seems pertinent to focus on the role that H might

play in weakening interfaces. In particular, it would seem unnecessarily limiting to confine attention to effects of H on the lattice per se, whereas this can emerge in the present treatment as the special case of an interface having no surface excesses of thermodynamic properties.

Attention will be directed to both the initiation and spread of fracture on interfaces. In the simplest case of elastic-brittle behavior, the inception of interface separation requires the attainment of a stress of magnitude equal to σ_m , where σ_m is the maximum cohesive stress. The spread of such a fracture, in crack-like form under nominal stresses much lower than σ_m , is governed instead by the cleavage energy γ_c (that is, by the total area beneath the cohesive stress vs. separation distance curve).

Both σ_m and γ_c are reduced, in general, by the presence of H. Also, this reduction is the source of diffusion controlled kinetic effects, which may operate over different size scales. In particular, suppose that the interface contains, before stressing, a surface excess Γ_i of H, at equilibrium under a potential μ_i for H. We may then define two limiting cases:

(i) Slow separation at constant potential, $\mu = \mu_i$; in this case the surface excess Γ will not, in general, remain constant during separation. Diffusion has adequate time for full equilibration of the mobile species (H) and, as will be shown, σ_m and γ_c then attain their lowest possible values, notably by comparison to the other limiting case of:

(ii) Rapid separation at constant interfacial concentration excess, $\Gamma = \Gamma_i$. The corresponding μ , that would equilibrate an excess Γ_i at various stages of separation, will not, in general, remain constant; the rates in this limiting case are imagined to be so great that diffusion is ineffective, and higher values of σ_m and γ_c then result.

The kinetic issues are even more complicated during macroscopic crack growth, because this initiation and spreading of interfacial cracks will usually be taking place in plastically deforming regions that are subjected to very substantial triaxial stress levels. The large stresses tend to augment the average local H concentration, by depressing the local μ and causing diffusion from less highly stressed regions. On the other hand the dislocations that are generated by the plastic flow cause the trapping of H , but may also transport H and perhaps, when densely piled up or annihilated with one another, cause local non-equilibrium concentrations.

Thus the effective values of Γ_i and μ_i that an interface sees, when it enters the separation process, may differ markedly from the nominal Γ and μ corresponding to similar interfaces at great distances from the macroscopic crack tip.

Thus, in addition to the small scale kinetics, implied by the differences between slow and rapid separation above, there are further kinetic factors emanating from the macroscopic crack tip plastic region and associated stress field, and these may be rate controlling. In such cases, brittle

interfacial failure should take place under the "slow" conditions (μ constant during separation), but with the corresponding μ value being that for adjacent material in the plastic zone; this μ value will be depressed in general, relative to that of points remote from the crack tip.

THERMODYNAMICS OF A SEPARATING INTERFACE

Consider an interface of area A separating two materials which, for simplicity, will be taken as having identical properties. The interface is subject to a normal stress σ , positive in tension. Also, the excess (in the Gibbs sense of surface excesses) displacement of one side of the interface relative to the other is δ ; δ will be called the "interfacial separation". The surface excess concentration of the mobile species (H) is Γ , and it is equilibrated by a potential μ . Hence, if u and s are the respective surface excess densities of internal energy and entropy, the Gibbs relation is

$$\sigma d\delta + \mu d\Gamma - s dT = df , \quad f = u - Ts \quad (1)$$

Here T is temperature and f is the excess Helmholtz function per unit area of surface.

Of specific interest are the equations of state

$$\sigma = \bar{\sigma}(\delta, \mu, T) , \quad \sigma = \hat{\sigma}(\delta, \Gamma, T) \quad (2)$$

because, from these, one can derive the maximum strength σ_m

and cleavage separation energy γ_c ($= \int \sigma d\delta$) for the two limiting cases, enumerated above, of : (i) separation at constant μ , and (ii) separation at constant Γ . In both cases it has been assumed tacitly that $T = \text{constant}$.

The following reciprocal relations, from (1), are useful in obtaining the form of these relations:

$$\left(\frac{\partial \sigma}{\partial \mu} \right)_{\delta, T} = - \left(\frac{\partial \Gamma}{\partial \delta} \right)_{\mu, T} \quad (3)$$

$$\left(\frac{\partial \sigma}{\partial \Gamma} \right)_{\delta, T} = \left(\frac{\partial \mu}{\partial \delta} \right)_{\Gamma, T} \quad (4)$$

CLEAVAGE SEPARATION ENERGIES AND GENERALIZATION OF THE GIBBS ADSORPTION THEOREM

In keeping with the earlier notation, let $\bar{\gamma}_c(\mu)$ be the cleavage separation energy at constant μ , and let $\hat{\gamma}_c(\Gamma)$ be the separation energy at constant Γ . Then (repressing T henceforth as an explicitly listed variable):

$$\bar{\gamma}_c(\mu) = \int_{\bar{\delta}_0(\mu)}^{\infty} \bar{\sigma}(\delta, \mu) d\delta \quad (5)$$

$$\hat{\gamma}_c(\Gamma) = \int_{\hat{\delta}_0(\Gamma)}^{\infty} \hat{\sigma}(\delta, \Gamma) d\delta \quad (6)$$

where the quantities $\bar{\delta}_0(\mu)$, $\hat{\delta}_0(\Gamma)$ are the ambient interfacial separations, corresponding to the given μ or Γ , when the stress $\sigma = 0$. From (1),

$$\bar{\gamma}_c(\mu) = \bar{f}_\infty(\mu) - \bar{f}_0(\mu) - \mu[\Gamma_\infty(\mu) - \Gamma_0(\mu)] \quad (7)$$

$$\hat{\gamma}_c(\Gamma) = \hat{f}_\infty(\Gamma) - \hat{f}_0(\Gamma) \quad (8)$$

where now: the subscripts "0" and "∞" refer respectively to the unstressed coherent interface, and to the infinitely separated interface, which then has the form of two free surfaces; $\bar{f}_{0,\infty}(\mu)$, $\hat{f}_{0,\infty}(\Gamma)$ denote the free energy in these limiting cases, regarded either as a function of μ or Γ ; and the functions $\Gamma_0(\mu)$, $\Gamma_\infty(\mu)$ are defined from the equilibrium μ vs. Γ relation; imagined to take the form $\Gamma = \Gamma_0(\mu)$ for a coherent interface and $\Gamma = \Gamma_\infty(\mu)$ for the infinitely separated free surfaces. Note that f_∞ and Γ_∞ are twice the free energy and excess concentration of a single free surface. Note also that when there is no interface, but instead a coherent lattice is being separated, we need only set $f_0 = 0$, $\Gamma_0 = 0$ in (7,8).

A generalization of the Gibbs adsorption theorem is obtained from the definition (5) for $\bar{\gamma}_c(\mu)$ and from the reciprocal relation (3):

$$\frac{d\gamma_c(\mu)}{d\mu} = - \int_{\delta_0(\mu)}^{\infty} \left[\frac{\partial \Gamma}{\partial \delta} \right]_\mu d\delta = -[\Gamma_\infty(\mu) - \Gamma_0(\mu)] \quad (9)$$

Thus the variation with μ or, equivalently, equilibrating H_2 gas pressure, of the equilibrium (i.e., constant μ) cleavage separation energy is determined directly by the difference in total surface excess between the two free surfaces and the unstressed interface. Petch has applied the analogous equation,

when $\Gamma_0 = 0$, to an explanation of H embrittlement.

An analogous equation follows from (6,4):

$$\frac{d\hat{\gamma}_c(\Gamma)}{d\Gamma} = \int_{\hat{\delta}_0(\Gamma)}^{\infty} \left(\frac{\partial \mu}{\partial \delta} \right)_{\Gamma} d\delta = \mu_{\infty}(\Gamma) - \mu_0(\Gamma) \quad (10)$$

where, in accord with earlier equations, $\mu = \mu_{\infty}(\Gamma)$ are equations of state for the completely separated free surfaces and for the unstressed coherent interface, respectively.

DIFFERENCE BETWEEN "SLOW" AND "FAST" SEPARATION ENERGIES

Consider now an unstressed coherent interface at equilibrium with $\mu = \mu_i$, $\Gamma = \Gamma_i$ [in terms of the functions of (9,10), $\Gamma_i = \Gamma_0(\mu_i)$, $\mu_i = \mu_0(\Gamma_i)$]. If it is separated slowly the cleavage energy is $\bar{\gamma}_c(\mu_i)$; if it is separated rapidly, by comparison to transport times for the mobile species, the separation energy is $\hat{\gamma}_c(\Gamma_i)$. In order to set limits to the range of applied loadings on a cracked interface, over which transport of the mobile species could govern the kinetics of crack growth, it is desirable to calculate the difference:

$$\hat{\gamma}_c(\Gamma_i) - \bar{\gamma}_c(\mu_i) \quad .$$

To obtain a formula for this, it is simplest to use Eqs. (7,8), noting that by the supposed initial equilibrium relation between μ_i and Γ_i ,

$$\bar{f}_0(\mu_i) = \hat{f}_0(\Gamma_i) \text{ and } \Gamma_i = \Gamma_0(\mu_i).$$

Thus, from (7,8),

$$\hat{\gamma}_c(\Gamma_i) - \bar{\gamma}_c(\mu_i) = \hat{f}_\infty(\Gamma_i) - \bar{f}_\infty(\mu_i) + \mu_i [\Gamma_\infty(\mu_i) - \Gamma_i] . \quad (11)$$

Now we note that since $\Gamma_\infty(\mu_j)$ is the excess concentration corresponding to an infinitely separated surface at potential μ_i ,

$$\bar{f}_\infty(\mu_i) = \hat{f}_\infty[\Gamma_\infty(\mu_i)] ,$$

so that (11) becomes

$$\hat{\gamma}_c(\Gamma_i) - \bar{\gamma}_c(\mu_i) = - \int_{\Gamma_i}^{\Gamma_\infty(\mu_i)} \frac{df_\infty(\Gamma)}{d\Gamma} d\Gamma + \mu_i [\Gamma_\infty(\mu_i) - \Gamma_i] .$$

But, from (1) when appropriately specialized,

$$\mu_\infty(\Gamma) = \frac{df_\infty(\Gamma)}{d\Gamma} ,$$

and the difference between the "fast" and "slow" separation energy is

$$\hat{\gamma}_c(\Gamma_i) - \bar{\gamma}_c(\mu_i) = \int_{\Gamma_i}^{\Gamma_\infty(\mu_i)} [\mu_i - \mu_\infty(\Gamma)] d\Gamma . \quad (12)$$

It is somewhat more revealing to write this in terms of

$$\Gamma_f \equiv \Gamma_\infty(\mu_i) , \quad (13)$$

noting that Γ_f is the final surface excess on two free surfaces that have been separated at slow, constant $\mu (= \mu_i)$ conditions. Then the difference in separation energies becomes, with obvious differences in notation,

$$(\gamma_c)_{\text{fast}} - (\gamma_c)_{\text{slow}} = \int_{\Gamma_i}^{\Gamma_f} [\mu_\infty(\Gamma_f) - \mu_\infty(\Gamma)] d\Gamma . \quad (14)$$

It is noteworthy that the integrand involves only the equation of state $\mu = \mu_\infty(\Gamma)$ for infinitely separated free surfaces, although the properties of the initial interface do enter implicitly, in determining the relation between the upper and lower limits of integration. But independently of these limits, we conclude that whenever Γ is a non-decreasing function of μ ,

$$(\gamma_c)_{\text{fast}} \geq (\gamma_c)_{\text{slow}} . \quad (15)$$

In this sense, the kinetics of transport of a mobile species to or from a surface should serve to stabilize an interfacial crack against rapid growth, in the sense that more fracture energy is required for fast (but still quasi-static) crack spreading than for slow.

It may be noted that from Eq. (9), the cleavage energy for slow (constant μ) interfacial separation could, presumably, be either an increasing or decreasing function of μ . Indeed, in the notation of (14), the variation with μ becomes

$$\frac{d\bar{\gamma}_c(\mu_i)}{d\mu_i} = -[\Gamma_f - \Gamma_i] \quad (16)$$

and there seems to be no a priori reason for which Γ_f must exceed Γ_i . But regardless of whether it does or not, the inequality (15) remains valid, in strict form whenever Γ_i and Γ_f differ.

EFFECT OF A MOBILE SPECIES ON THE MAXIMUM COHESIVE STRESS

Consider the equation of state $\sigma = \bar{\sigma}(\delta, \mu)$ and define $\bar{\delta}_m(\mu)$ as the value of δ at maximum stress. That is,

$$\frac{\partial \bar{\sigma}(\delta, \mu)}{\partial \delta} = 0 \quad \text{when } \delta = \bar{\delta}_m(\mu) , \quad (17)$$

and the maximum cohesive stress in "slow" separation, at constant μ , is

$$\bar{\sigma}_m(\mu) \equiv \bar{\sigma}[\bar{\delta}_m(\mu), \mu] . \quad (18)$$

Similarly, one may start with equation of state $\sigma = \hat{\sigma}(\delta, \Gamma)$, defining $\hat{\delta}_m(\Gamma)$

$$\frac{\partial \hat{\sigma}(\delta, \Gamma)}{\partial \delta} = 0 \quad \text{when } \delta = \hat{\delta}_m(\Gamma) , \quad (19)$$

and the maximum cohesive stress in "fast" separation, at constant Γ , is

$$\hat{\sigma}_m(\Gamma) \equiv \hat{\sigma}[\hat{\delta}_m(\Gamma), \Gamma] . \quad (20)$$

Thus, from (18), with the use of (17), one may write

$$\frac{d\bar{\sigma}_m(\mu)}{d\mu} = \left[\frac{\partial \bar{\sigma}(\delta, \mu)}{\partial \mu} \right]_{\delta=\bar{\delta}_m(\mu)} .$$

But now the reciprocal relation (3) may be used, and one derives

$$\frac{d\bar{\sigma}_m(\mu)}{d\mu} = - \left[\frac{\partial \Gamma(\delta, \mu)}{\partial \delta} \right]_{\delta=\bar{\delta}_m(\mu)} \quad (21)$$

Hence, in "slow" separation, at constant potential, μ , the increase in surface concentration per unit increase in sepa-

ration, when evaluated for the separation corresponding to maximum strength, gives the rate at which the maximum strength $\bar{\sigma}_m(\mu)$ decreases with increasing potential, μ .

In an analogous way, one obtains

$$\frac{d\bar{\sigma}_m(\mu)}{d\Gamma} = \left[\frac{\partial \mu(\delta, \Gamma)}{\partial \delta} \right]_{\delta=\hat{\delta}_m(\Gamma)} \quad (22)$$

SPECIAL DILUTE CONCENTRATION FUNCTIONAL FORMS

So far our considerations are general and no specific assumptions have been made as to the form of the equations of state for the interface. Let us now consider the dilute concentration form for the surface excess of free energy:

$$f = u - TS = F_0(\delta) + \Gamma\beta(\delta) - TS_{\text{mix}}(\delta, \Gamma) \quad (23)$$

Here the explicit listing of T as a variable is again avoided. The first term $F_0(\delta)$ is the energy in the absence of a mobile species. In the last, $S_{\text{mix}}(\delta, \Gamma)$ is the entropy of mixing, so that for a surface of area A

$$AS_{\text{mix}} = k \log \left[\frac{[AN(\delta)]!}{[A\Gamma]! [AN(\delta) - A\Gamma]!} \right], \quad (24)$$

where k is the Boltzmann factor and $N(\delta)$ is the maximum possible density of excess solute sites per unit area of the interface, when it has been separated by the amount δ . The occupation of sites is approximated as being statistically independent. The term $\beta(\delta)$ is the increase in the total internal energy, minus T times the vibrational part of the entropy, when one solute

atom is added to a surface of fixed separation, δ . We refer to β as the "binding energy"; it differs from the corresponding energy change when an atom of solute is added to the interface while the stress, σ , rather than separation, is held fixed.

By using the Stirling approximation in the usual way in (24), the surface excess of free energy becomes:

$$f = F_0(\delta) + \Gamma\beta(\delta) - kT \left\{ \Gamma \log \left[\frac{N(\delta) - \Gamma}{\Gamma} \right] - N(\delta) \log \left[\frac{N(\delta) - \Gamma}{N(\delta)} \right] \right\} \quad (25)$$

From this and the Gibbs equation, (1),

$$\mu = \beta(\delta) + kT \log \left[\frac{\Gamma}{N(\delta) - \Gamma} \right]. \quad (26)$$

Further, we obtain σ in the form $\hat{\sigma}(\delta, \Gamma)$ as

$$\sigma = \hat{\sigma}(\delta, \Gamma) = \sigma_0(\delta) + \Gamma\beta'(\delta) - kT N'(\delta) \log \left[\frac{N(\delta)}{N(\delta) - \Gamma} \right], \quad (27)$$

where $\sigma_0(\delta) \equiv F'_0(\delta)$ and the "prime" means derivative with respect to δ .

Alternatively, we may solve (26) to obtain

$$\Gamma = N(\delta) \frac{x(\delta, \mu)}{1+x(\delta, \mu)}, \text{ where } x \equiv \exp \left[\frac{\mu - \beta(\delta)}{kT} \right], \quad (28)$$

By substituting this for Γ in (27), we obtain

$$\sigma = \bar{\sigma}(\delta, \mu) = \sigma_0(\delta) + N(\delta)\beta'(\delta) \frac{x(\delta, \mu)}{1+x(\delta, \mu)} - kT N'(\delta) \log[1+x(\delta, \mu)].$$

These results may be used to compute the various parameters of interest, e.g., $\bar{\gamma}_c(\mu)$, $\bar{\sigma}_m(\mu)$, $\hat{\gamma}_c(\Gamma)$, $\hat{\sigma}_m(\Gamma)$, etc. that appear in the earlier discussion, provided that functional forms of $\sigma_0(\delta)$, $\beta(\delta)$, and $N(\delta)$ are given. None of these are

known very accurately, but schematic variations are sketched in the figure.

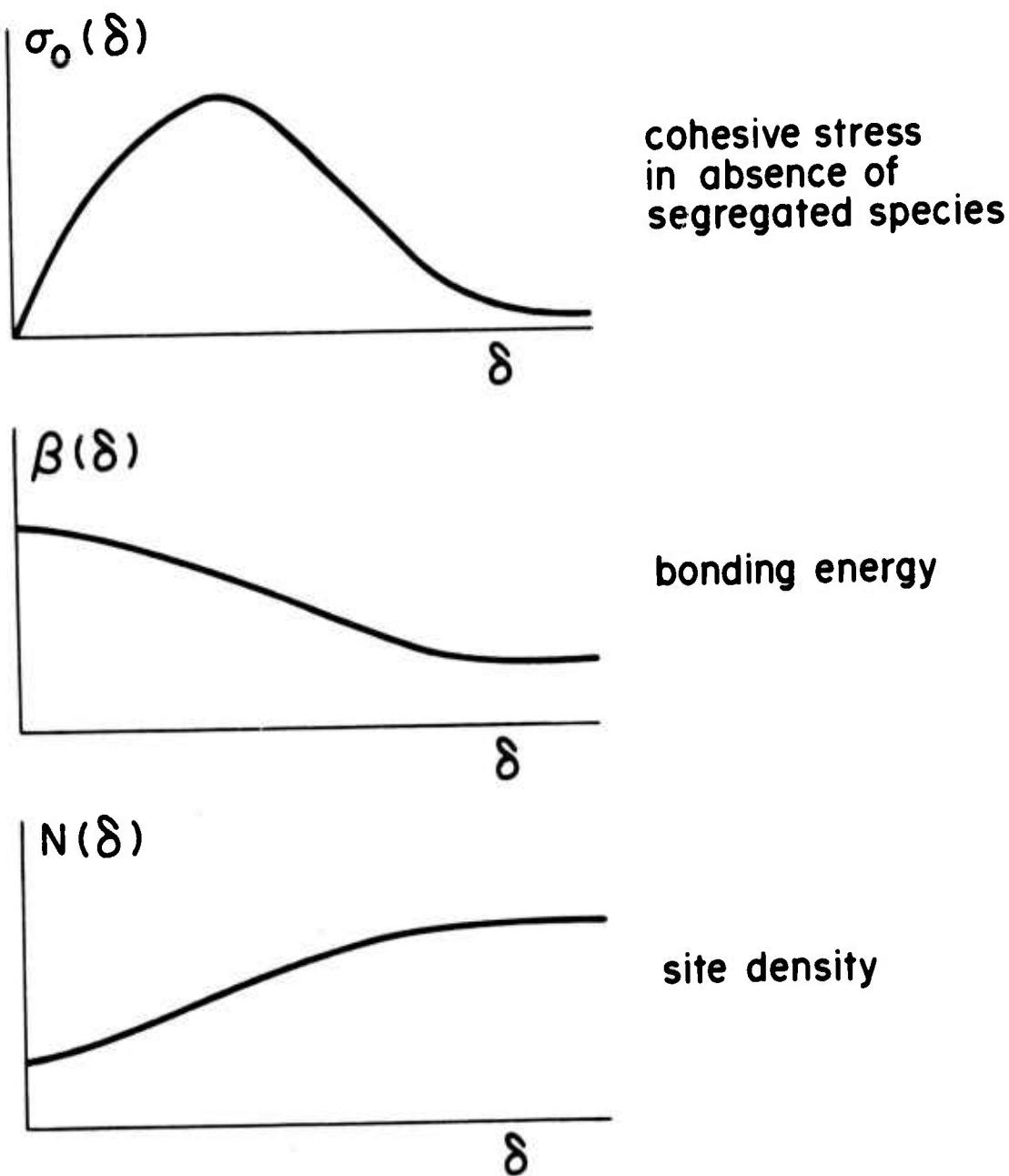
CONCLUDING DISCUSSION

This is an interim report on studies at the July 1975 MRC meeting. It is clear that further study and modeling of the functional forms, in light of evidence from experimental measurement, is needed to carry the work to numerical estimates of the various separation parameters.

The discussion here is phrased in terms of "slow" and "fast" response, with no explicit account of the kinetics and the time scale. However, the formulation of equations such as (26, 27), in the form

$$\sigma = \hat{\sigma}(\delta, \Gamma), \quad \mu = \hat{\mu}(\delta, I)$$

is a first step. A coupled elasticity/diffusion calculation is then implied, in the case of crack growth, for example. The distribution of σ on the interface can be expressed in terms of the remotely applied stress and the (unknown) distribution of δ along the interface. This must give a σ that is consistent with the first equation of state above, thus coupling the stress distribution to that of Γ , also unknown. Finally, gradients in μ drive diffusive fluxes, whose divergence in the adjoining material or along the interface must be consistent with the rate of increase of solute concentration. Since μ depends on δ , the diffusion field is coupled likewise to the elasticity field.



FIGURE

But apart from the kinetics of growth, when growth is indeed possible, there is considerable interest in the limiting threshold below which no crack growth seems to occur. Here the pertinent parameters are $\bar{\gamma}_c(\mu)$ and $\bar{\sigma}_m(\mu)$. But their application to the determination of threshold conditions may not be simple, especially in the regime of HF in steels where the effect of H is simply to enable, at lower K levels, the progression of the normal sort of ductile rupture mechanisms by the earlier nucleation of voids than would normally be the case.

Even when the fracture path, just above threshold, would be a brittle, continuous intergranular path (or, perhaps exceptionally, a lattice plane path), the determination of threshold conditions is far from straightforward. This is because the fracture must, in general, be expected to be accompanied by some local dislocation motion and generation, which will screen the crack tip and result in a macroscopic energy release rate far in excess of $\bar{\gamma}_c(\mu)$.

ACKNOWLEDGEMENTS

I am grateful to R. Thomson for helpful discussions. This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

MATERIALS/DESIGN

A. G. Evans

DESIGN PROCEDURES

A design procedure, as it evolved from group discussions, which entails the optimum interaction between all recognized participants in the design process is summarized in Fig. 1.

The initial design input is determined by the application, which essentially specifies a series of operating parameters, listed in Table I. Based on these parameters, the designers can generate a semi-quantitative initial design, utilizing past experience, approximate calculations, etc., and the materials engineer can establish an initial materials inventory (which, for example, recognizes the inadmissibility of certain materials due to environmental constraints). The latter can be achieved if an adequate data base, which includes at least the parameters listed in Table II, is available for each material. At this stage a comparison of the initial design with the material properties provides a basis for estimating the feasibility of the component design, based on available materials.* Then, if manufacturing viability is

*An estimate of the potential for developing the superior material can also be incorporated, if available materials appear inadequate.

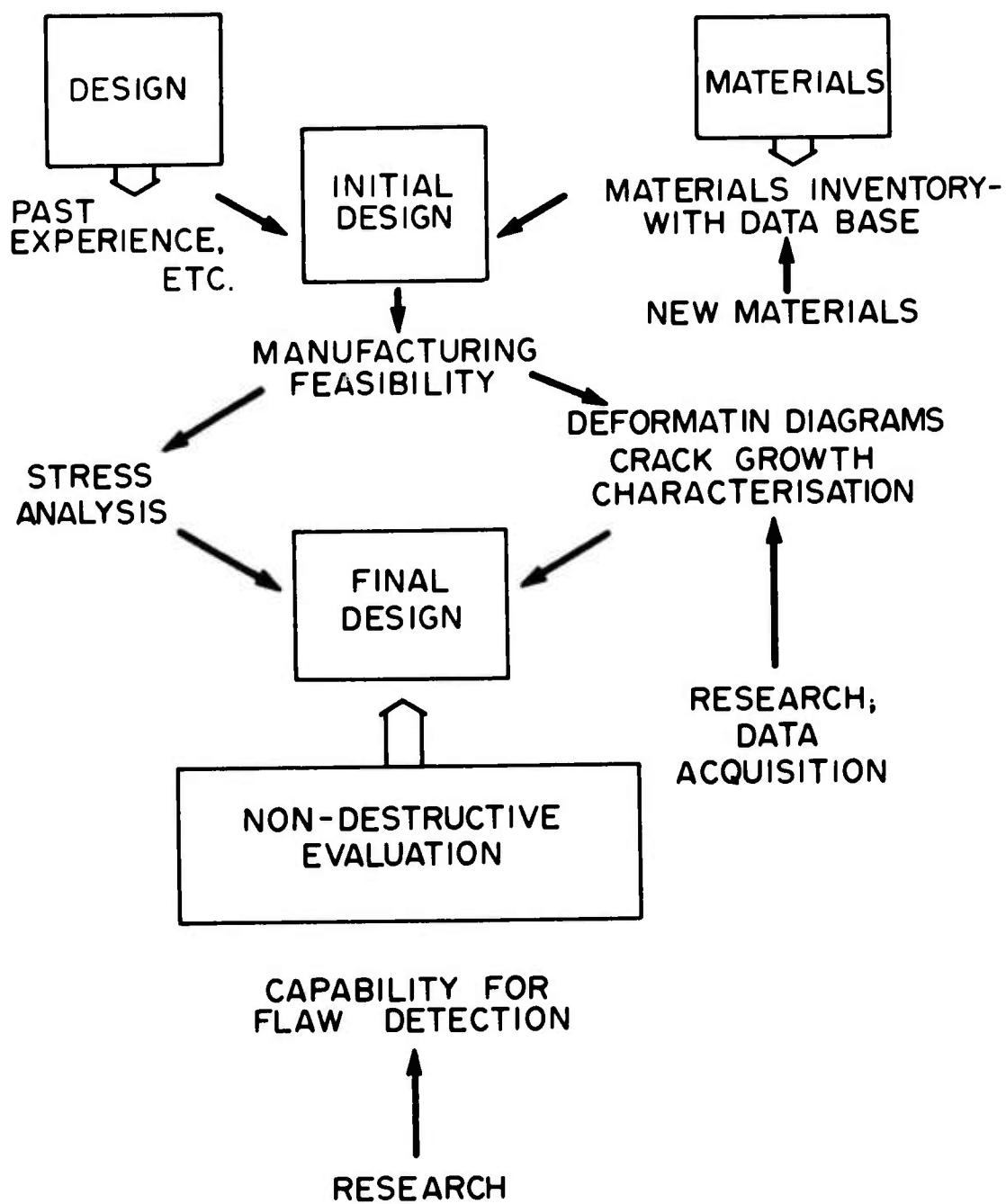


Figure 1. A schematic of a typical design procedure.

TABLE I.

OPERATING PARAMETERS
ENVIRONMENT
TEMPERATURE
VIBRATION SPECTRUM
LIFETIME
FAILURE PROBABILITY
CONFIDENCE LIMIT

TABLE II.

DATA BASE (WITH CONFIDENCE LIMITS) AS FUNCTION OF TEMPERATURE
ELASTIC CONSTANTS (E, v)
THERMAL EXPANSION COEFFICIENT (α)
STRESS/STRAIN RELATIONS (σ, ε)
FRACTURE TOUGHNESS (K _C)
STRESS CORROSION SUSCEPTIBILITY (K _{scc} , environment)
FATIGUE SUSCEPTIBILITY (da/dN vs. ΔK)
CORROSION FATIGUE SUSCEPTIBILITY (da/dN vs. ΔK, K̄, λ)

also established, more quantitative design procedures can be initiated.

The final design phase commences with a stress analysis (ideally finite element) which estimates the stress levels in each critical element within the component, for the anticipated service load spectrum. Then, the non-destructive evaluation group evaluates flaw detectability, in relation to the geometry of the part,* and specifies a maximum flaw size that can be reasonably detected with available (or soon to be available) techniques. The materials engineers then supply the deformation and crack growth data that fully characterize the mechanical performance of the candidate materials, under the service condition. Finally, the deformation and the crack growth (based on the NDE initial flaw size specification) are computed (again by finite element techniques) for a few of the more promising materials, to estimate the dimensional changes and crack extension that should occur during the lifetime of the component. This procedure provides a quantitative measure of the acceptability of each material for this application, based on the initial design. Possible design modifications can be explored at this stage, if the performance of available materials with the initial design appears unsatisfactory.

*The NDE group should provide an important input in the final design phase, to ensure optimum flaw detectability, i.e., averting inaccessibility, etc., within the constraints of the system.

DATA PRESENTATION

The techniques for obtaining the basic material constants, the deformation characteristics and the crack growth behavior (under linear elastic conditions) are available. There are thus no technical barriers to the acquisition of this data; but data accumulation under all of the service conditions (quasi-static, mechanical fatigue and thermal fatigue conditions) is expensive and there are economic constraints. This problem will be addressed in more detail in a later section.

Conversely, techniques for characterizing crack growth under non-linear (plastic) conditions are not presently available, except for the onset condition of crack growth. This limitation will also be discussed in a later section.

The presentation of available data, in a form that is readily comprehended by designers, is an important part of the technical communication between materials engineers and designers. The deformation or crack growth behavior cannot, of course, be represented on single diagrams; but first order approximations, which could be invaluable for initial design, can be provided in a series of diagrams.

A design diagram which adequately represents steady state uniaxial deformation for conditions of constant micro-structure has been proposed by Ashby. A schematic of a typical diagram is shown in Fig. 2. These diagrams indicate that there are regimes of behavior (A,B,C,D) where the functional

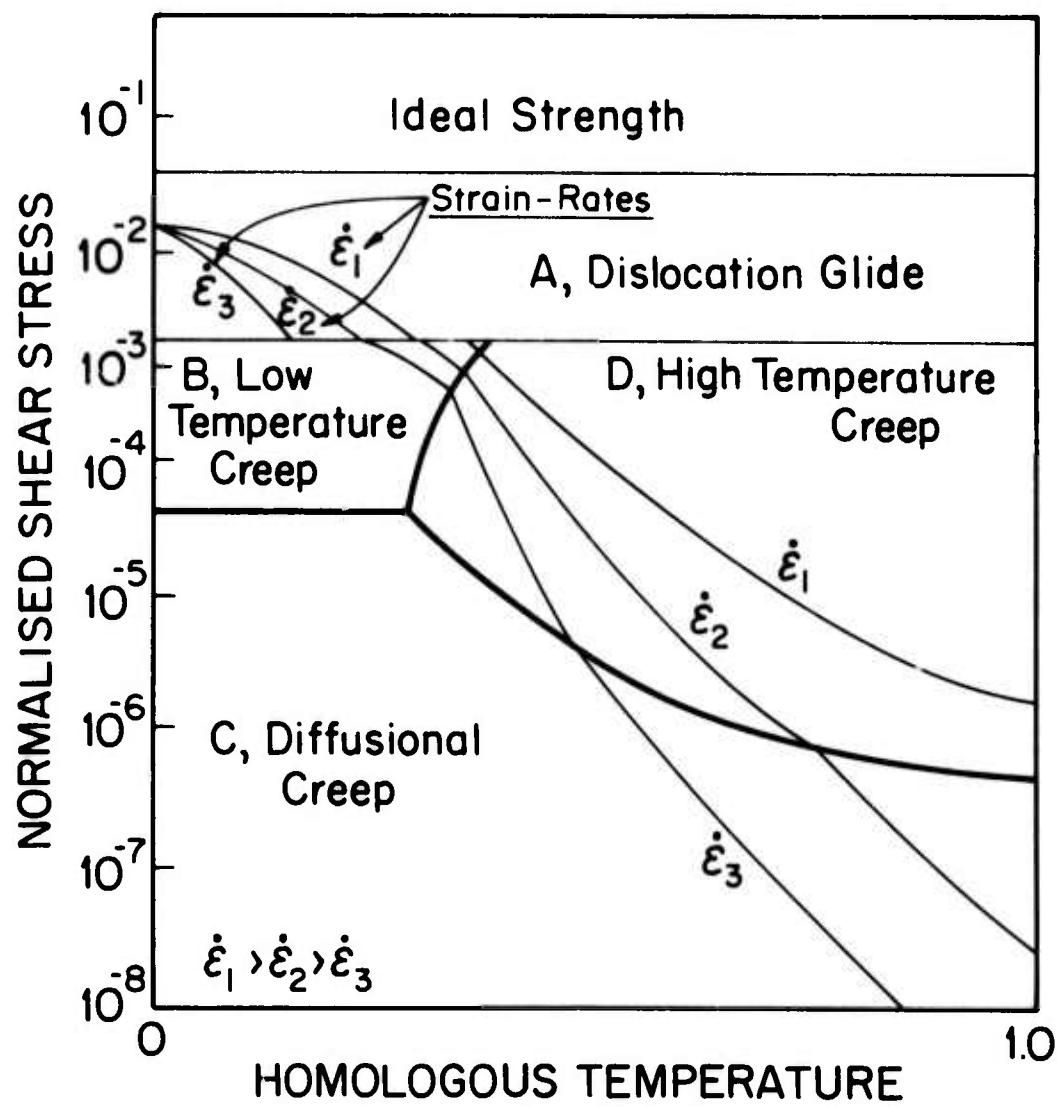


Figure 2. A schematic of a creep diagram showing four rate controlling mechanisms over the regimes of stress and temperature of practical interest.

forms of the deformation equations are different; and provide information about the steady state deformation rate within each regime, for particular values of the stress and temperature. The limitations of the diagrams are, of course, their exclusion of primary (or transient) deformation, their lack of recognition of time dependent microstructural changes, and their restriction to uniaxial deformation. These omissions might be rectified by supplying information on transient behavior, microstructural changes and multiaxial response in a series of separate diagrams or tables.

The crack growth behavior for specific stress modes can also be represented by diagrams. A useful representation for design purposes, plots the lifetime as a function of stress (or stress amplitude) for a series of initial flaw sizes (Fig. 3). Hence, by knowing the flaw detection capability for a component the "stress allowable" for the required lifetime can be read off the diagram. For ceramic materials the diagrams can completely represent the fracture behavior (for a specified environment), by supplying tables or graphs of multiplying constants for the lifetime axis - which allow for variations in temperature, concentration of stress corrosive species, stress amplitude and frequency, etc. For metallic (and probably polymeric) materials the diagrams are not as "all embracing". The primary problems are the frequency and mean stress dependence of the lifetime for corrosion fatigue conditions and the history (e.g., overload) effects observed in fatigue. These

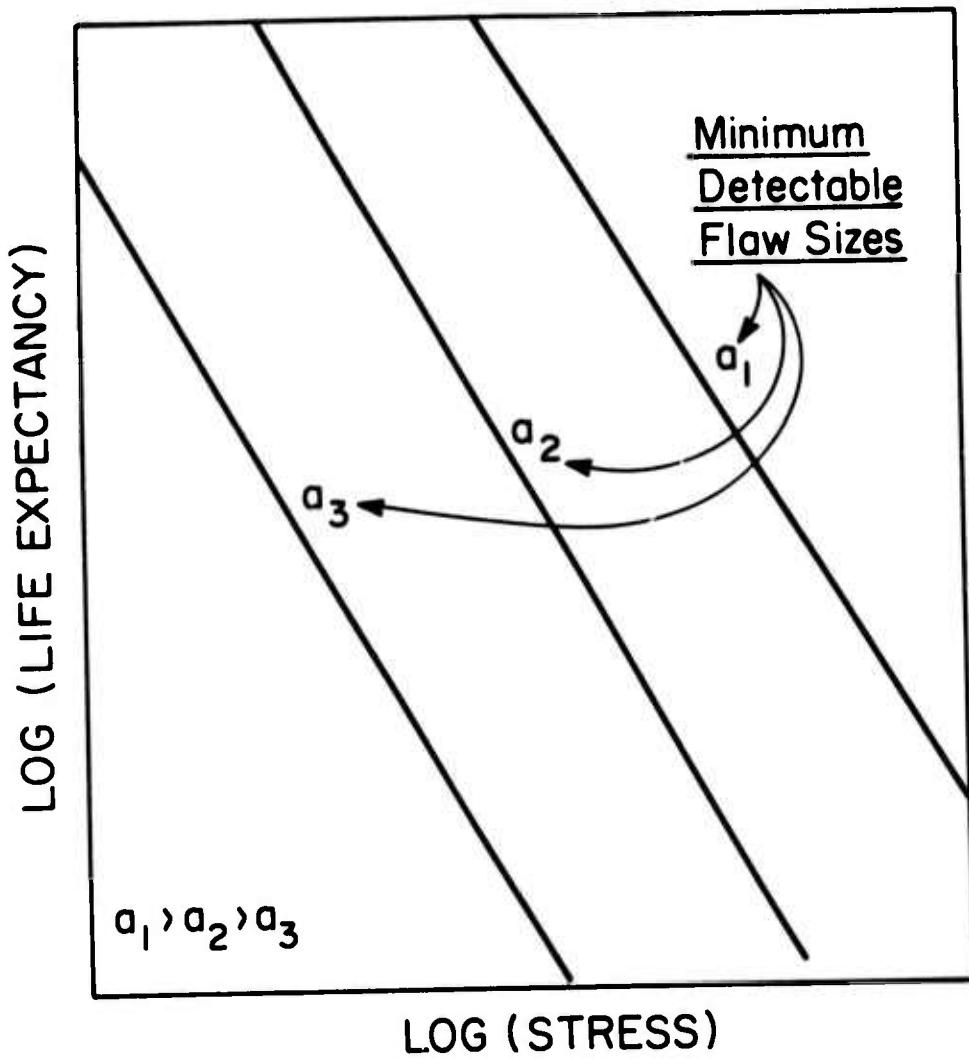


Figure 3. A schematic of a failure diagram derived from fracture mechanics data for specified flaw detection capabilities.

effects observed in fatigue. These effects must presently be evaluated empirically and they are difficult to incorporate effectively in failure diagrams.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

RESEARCH ON MATERIAL CHARACTERIZATION FOR
STRUCTURAL RELIABILITY AND DESIGN

J. R. Rice

ABSTRACT

As an outgrowth of MRC discussions on the Design/
Materials Interface, some areas of materials research, mostly
in fracture mechanics, are identified which would seem to have
direct relevance to needs in structural reliability and design.

RESEARCH ON MATERIAL CHARACTERIZATION FOR
STRUCTURAL RELIABILITY AND DESIGN

J. R. Rice

Discussions on the Design/Materials Interface at the July 1975 MRC meeting focused, in part, on the extent to which recent developments in material characterization, especially those of fracture mechanics, are used in design. Evans (in this volume) has summarized the current situation, showing the manner in which established concepts on crack growth can serve as the basis for design estimates of admissible stress levels and service lifetimes.

However, it also emerged from the discussion that present knowledge of materials characterization, both in fracture mechanics and constitutive descriptions, is strikingly inadequate to provide guidance to designers on issues which they must frequently consider. The purpose of this report is, therefore, to identify some areas of basic material characterization in need of further development, with special reference to those topics for which success in research could be expected to have a direct bearing on structural reliability and design.

The outline to follow is certainly not exhaustive, but reflects instead the specific problems raised in discussions by those attending the meeting, and perhaps also the interpretations of the writer.

ANALYTICAL IMPLEMENTATION OF FLAW GROWTH CRITERIA

Basic fracture mechanics is now well developed in terms of the phrasing of fatigue crack growth rates in terms of ΔK and K_{mean} , of the characterization of catastrophic crack growth in terms of K_{Ic} for nominally elastic surroundings, and, although it is still somewhat a research area, of the characterization of the onset of crack growth in terms of J_{Ic} even well into the general yielding range.

The application of these flaw-growth concepts in structural design is, however, limited by the geometric complexity of crack starting sites (e.g., boltholes, nozzle intersections) and/or of the three-dimensional crack configurations that result. Specifically, very difficult three-dimensional stress analyses, extending beyond what is normally done in design studies, are required to implement the growth criteria. Such analyses are now possible, utilizing advanced numerical techniques in finite element and/or boundary integral equation approaches.

Hence it would seem desirable to prepare, for design use, a short catalog of flaw growth analyses for representative geometries. One of these of interest for aircraft structures is the crack emanating from a rivet or bolthole. As indicative of what is desired for fatigue lifetime prediction, a suitable range of initial flaw shapes at the hole should be considered, and their three-dimensional growth reported for various forms of the relation between da/dN and ΔK . Here the N may refer to

the number of ground-air-ground cycles, during each of which a non-cyclic load program is applied, and ΔK is some representative K range during the program. Additional factors determining the form of the da/dN vs. ΔK relation are the ratio of mean K to ΔK and the environment; it is regretable that the influence of none of these is available on fundamental grounds, but suitable test data on pre-cracked laboratory specimens can be taken to provide the necessary relations.

Another type of calculation, of interest to pressure vessel technology, is the part-through wall crack in a plate or shell. This type of problem has already been studied extensively on the basis of elastic fracture mechanics, and results have been summarized in a form useful to design. However, the usual materials of construction (e.g., medium strength steels) require consideration of large scale plastic yielding at the flaw site. In such cases it would seem desirable to obtain a similar catalog of solutions relating the J integral to applied load, crack geometry, and plastic flow properties, and to summarize results for design purposes in terms of J_C .

The above example studies and many others can be completed successfully with the more advanced of current computing techniques. However, the cost of 3 dimensional analyses is high and often the requisite techniques for crack analysis are quite specialized and unavailable to those who are unable, for example, to tamper significantly with the software of general purpose analysis packages. Thus a continuing effort should be

supported on the pursuit of new computational methods for design implementation of fracture mechanics, the emphasis being on methods that: are applicable for general three-dimensional crack problems, are attractive from the standpoint of computer time, and, preferably, can readily be used by non-specialist stress analysts and designers so that actual structural geometries, flaw origins, and material responses are readily incorporated into reliability and lifetime predictions.

FATIGUE CRACK GROWTH MODELS

The crux of the elastic fracture mechanics approach to fatigue is that the rate of crack growth is dependent on the K variation sensed at the crack tip. This gives an extremely useful tool for translating laboratory measurements into predictions of flaw growth in structures. But the approach gives no hint as to the general functional form of the growth rate vs. K relation, and very extensive and costly laboratory testing is implied if a wide range of environmental, rate, mean load, and non-cyclic loading effects is to be considered in the structural application. Indeed, the last of these is particularly vexing, for it is known that a tensile overload during cyclic loading can greatly retard the subsequent cyclic growth rate, whereas compressive overloads have an accelerating effect.

The practical advantage of a rational model for fatigue crack growth is self-evident. This is, of course, hardly a new research area, but a number of recent developments suggest that

the time may be ripe for a more concerted research effort toward a predictive theory of fatigue crack growth.

These developments include: The experimental means to detect small amounts of crack extension, and thus to measure accurately closure near a crack tip and even, at sufficiently high rates, the growth in a single cycle; the general elucidation of the structure of plastic deformation and stress fields very near a crack tip, and also the availability of powerful analytical tools for further modeling of crack tip processes; and, most importantly, the presence of a reasonably comprehensive conceptual view of the fatigue crack growth process that seems to agree qualitatively with observed phenomena, and that has strong promise for translation into predictive models, based on a reasonably small amount of laboratory data.

To elaborate on the conceptual view, it seems that the fatigue crack growth Δa during a tensile K excursion, say of magnitude ΔK , is essentially equal to a fraction (roughly one-half) of the crack tip opening displacement, $\Delta \delta$, during that excursion, but with the following important, and often overriding qualifications:

- i) Only a fraction, ΔK_{eff} , of the total ΔK is actually effective in producing $\Delta \delta$ because of crack surface closure,
- ii) The actual fraction, say α , of the $\Delta \delta$ that results in crack growth is environment sensitive. Active environments may allow all of the plastic tip opening to result from slipping off along fresh surfaces at the tip so that $\alpha = 1$, whereas less

active environments may allow some of the slipping-off to emanate on previously slid systems making $\alpha < 1$. The fraction α will, in general, be dependent also on any other factors influencing processes of slip initiation at or near a surface, or factors such as stacking fault energy affecting the character of slip. It could also reflect, in some cases, partial rewelding upon closure.

iii) The plastic opening at the tip takes place in material which has been preconditioned by previous cycling and will, except perhaps at low ΔK in fairly pure materials, involve some stable ductile rupturing by void nucleation and coalescence or, exceptionally, by cleavage microfaceting. If δ_c is the tip opening displacement at the onset of rupture (vs. crack blunting or stretch zone formation) in monotonic loading, then an approximate measure of the contribution from stable rupturing is $\Delta\delta/\delta_c$, and some function $f(\Delta\delta/\delta_c)$ will give the amplification of the growth due to this effect, where $f(0) = 1$. The function f may, in fact have to contain a wider range of variables, especially to accommodate strongly stress-dependent rupture processes or environmental influences on the rupture mechanism. Also, it is possible that when the contribution from stable rupture is great, some environmental influences, reflected in α , may be annulled. Hence the approximate suggested form of an expression for crack growth in the load excursion is;

$$\Delta a = \frac{1}{2} \alpha \Delta \delta f(\Delta\delta/\delta_c, \dots)$$

where $f(\sigma, \dots) = 1$, and where $\Delta\delta$ is related to the effective K range by

$$\Delta\delta \approx \frac{1}{4} (\Delta K_{\text{eff}})^2 / E \sigma_0$$

where σ_0 is a representative yield or flow stress level in the cyclic plastic zone. It will depend strongly on cyclic flow response.

Essentially, in planning a research program, it would seem a job for elastic-plastic continuum mechanics to deduce the relation of ΔK_{eff} , and hence $\Delta\delta$, to the imposed ΔK , in terms of crack closure as governed by the previous ΔK levels (e.g., overloads, residual stress fields, etc.) and to the surface asperities from prior load variations. Additional contributions from surface chemistry and dislocation theory would be necessary for determining α , and knowledge of ductile rupture mechanisms in combination with plasticity theory should help in formulating the function f .

There are, of course, many more features to be considered if a fully predictive and microscale based theory of fatigue is to be developed. But the emphasis here is not on full prediction, but rather on the development of rational procedures for estimating, from limited laboratory measurements, flaw growth in structural configurations subjected to general loadings.

FRACTURE WITH LARGE SCALE PLASTIC YIELDING

The utility of most of current fracture mechanics theory in design is severely limited by its restriction to nominally elastic conditions of the flawed structural element at failure. Such conditions, while normally met in fatigue, are met for fracture under monotonic loading only in structural alloys that suitably combine brittleness and largeness of size for the cracked element. Fracture in alloys of medium or lower strength normally involves, except in large structural sizes, general yielding over the precracked section. Further, even if the size of the structure, or of the critical structural element, allows an elastic fracture mechanics characterization in service, in terms of a K_C or K_{IC} , small pre-cracked laboratory specimens of the same material may not fail until loaded well into the general yielding range. The cost of testing full-sized elastic fracture mechanics specimens is staggering, and hence it is important to be able to predict the crack growth behavior of large and, possibly, nominally elastic structural elements from small generally yielded laboratory specimens.

These observations demonstrate the relevance to design needs of further developing elastic-plastic fracture mechanics. The area has received a strong impetus in recent years from work on characterizing the onset of plane strain crack growth, under monotonic loading, in terms of the J integral and related features of crack tip plastic fields. But conditions for the onset of growth are, for the more ductile alloys, far removed

from those for a terminal running crack instability. Indeed, the stable crack growth that results, generally, under further load increase and the conditions for its ultimate loss of stability are little understood. Still, the capabilities now available on elastic-plastic modeling, as well as the promising work on the onset of growth under fully plastic conditions, would seem to make this an area that is ripe for further development. The most promising approach would seem to be that of elastic-plastic continuum mechanics in conjunction with metallurgical observations and modeling of microscale rupture processes during stable crack growth. Of course, for design implementation one hopes to evolve procedures that require no detailed microscale modeling, but that can instead be implemented on the basis of macroscale experiments with precracked specimens. Still, it remains an open question as to what extent a theory of stable crack growth can be divorced from details, at some level, of the actual processes of separation.

Certainly, the current situation is that there is no universally agreed upon, or even widely agreed upon, criterion for continuing crack growth that can be fed into crack stress analysis program. Also, some ideas, now extant, relating to energy balance approaches for crack growth in elastic-plastic continua, are clearly defective. There is no reason to assume, a priori, that calculation of energy transfers at a continuum level can be decoupled, in the Griffith manner, from processes of energy transfer on the surface of separation. Indeed, this

decoupling was shown some years ago to lead to a paradox for elastic-plastic continua, in that no energy flux results in a continuum model with continuous crack advance, essentially because there is no singularity in recoverable energy (vs. total stress work) at the crack tip.

Some approaches, that are sufficiently simple from the standpoint of possible design implementation, include the model of a progressively separating cohesive zone, embedded in a surrounding elastic-plastic continuum, or perhaps the energy balance model, but based on discrete steps of crack advance, the step size being regarded as based (ultimately, if only empirically in the application) on microscale dimensions in ductile rupture. In addition to pursuing ideas such as these and others on modeling crack growth, attention should be directed also to the development of finite element techniques that are more suited to this class of problems. The advance of model cracks by one finite element length at a time has its deficiencies. What is needed are analysis methods which allow, at least in principle, continuous crack growth and the retention of a highly focused and detailed finite element mesh near the crack tip. This suggests research on elastic-plastic numerical approaches in which the mesh moves relative to the material.

FAILURE UNDER ELEVATED TEMPERATURE CREEP CONDITIONS

This is an area in which needs for material characterization are obvious, but it is also one for which, at present, there seems to be few ideas or approaches of a kind likely to bring significant advances, for design utilization, in the near future. The need is for both realistic constitutive descriptions of inelastic deformation, applicable to arbitrary stress histories, and descriptions of fracture, either by general cavitation or by the growth of a single dominant crack.

The prediction of failure by general cavitation can be lumped with constitutive description, but there is, at present, little in the way of theory or of combined stress experimentation directed toward developing the comprehensive understanding of material response necessary to rational stress analysis in design.

Also, for those circumstances in which a single macroscopic crack grows, it is clear that an elastic K value cannot be used as a correlating parameter, unless the material far removed from the crack tip is behaving in a linear elastic (or linear viscous) fashion. A generalization of the J integral has been proposed to correlate crack growth when elastic and transient plastic effects are unimportant and the material responds as if it were purely viscous. Such conditions will, however, be met only rarely in practice, and the need exists for a general mechanics-based approach to creep crack growth.

FUNDAMENTAL MODELING OF MICROSCALE FRACTURE PHENOMENA

The amounts of effort and funding devoted to fundamental, quantitative modeling of the physics and mechanics of fracture micromechanisms seem to represent a remarkably small fraction of the total activity in the fracture area. Certainly, it is optimistic to think that work in this area would have a short-term impact on design, but it is imprudent to think that its impact would be only on the design of materials (vs. the design of structures). For example, it should be possible on grounds other than experience and empiricism to state, in terms open to design use, the susceptibility of a given alloy to hydrogen attack, and to give quantitative measures of that susceptibility for differing levels of equilibrating potential and temperature. Similarly, it should be possible on fundamental grounds to decide which environment-alloy pairings will produce corrosive cracking and to quantify the effect in terms of K_{Iscc} or like parameters.

Apart from the direct relevance to questions such as these, the understanding of fracture events at the micro-level should provide an invaluable underpinning to the sorts of macroscopic rupture initiation and crack growth models that will find utilization in design. Indeed, formulations at the macro level are open to great arbitrariness in the choice of functional forms assumed to describe deformation and fracture phenomena. Thus, while it is probably unrealistic to expect, in the near future, the development of reliable material models

extending from microscale details to the macro level of design, it is essential that work at the macro-level be carried out in recognition of micro-level processes and models. This provides important guidelines on the choice of macroscopic functional forms, on the likely limitations of a given formulation, and on the suggestion of critical experiments. Such an understanding is of great importance when considering the response of structural materials over broad ranges of temperature, or of loading rates and patterns, or of surrounding environments.

ACKNOWLEDGEMENT

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHCl5-71-C-0253 with The University of Michigan.

THE MATERIALS/DESIGN INTERFACE

R. M. Thomson
M. Cohen

I. Introduction

On 7-8 July a small workshop was convened by the Materials Research Council at the request of ARPA to explore the way in which materials engineers function in the process of the design of new systems. The list of participants is appended (Attachment No. 1). In the background to the workshop is a NASA report, NASA TMX-68850 exploring some 231 structural failures; 48% of the failures were attributed to "poor design" and 30% to "improper fabrication". Thus, the suspicion emerges that perhaps materials people were not adequately involved during the initial system design, although only 1% of the failures were caused by "inadequate materials knowledge". (See Table I.)

Our focus at the workshop was narrowed to the mechanical failure aspects of the materials input to design, and specifically did not include such topics as corrosion, wear, new materials development, etc.

II. Nature of the Materials/Design Interface

1. Interrelation of Design, Materials and Manufacturing

During the life of a complex system such as a B-1 bomber or a nuclear power plant, the materials engineer is

TABLE I. FAILURE ANALYSIS MATRIX FOR 231 FAILURES^a
 Source: NASA TMX-68850

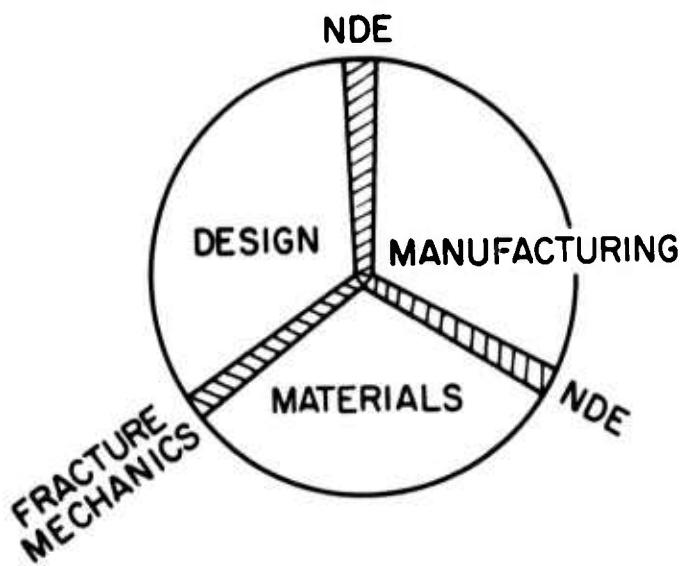
Cause	Mechanism					Cause related percent
	Fatigue	Corrosion	Hydrogen embrittlement	Oxidation	Other ^b	
Defective materials	--	1	--	--	8	4
Improper fabrication	18	17	5	--	35	30
Poor design	53	15	1	1	49	48
Improper use	3	8	1	--	11	9
Inadequate knowledge of service conditions	5	8	--	4	2	8
Inadequate materials knowledge	--	2 ^c	--	--	--	1

^aThe number of failures refer to structural types. In some cases several individual parts may be included in one type.

^bFailure by tensile overload or undermined.

^cVery serious failures with impact on five major program areas.

often a part of the materials development team in the early stages of the project; later, a member of the design team; later again, a member of the manufacturing quality control team; and later still, a member of the main finance or failure analysis team. During the design process, materials engineers function with design engineers and manufacturing engineers through NDE on the one hand, and the discipline of fracture mechanics on the other, as depicted in the following diagram.



The most important element which has allowed the materials engineer to contribute quantitatively to the design process is the development of fracture mechanics together with the ability to specify, for a particular material and service environment, the critical flaw size which must not be exceeded. However, fracture control is not yet a finished methodology by any means, and any advance by which reliability can be more exactly specified and quantified by the materials engineer

will contribute immediate cost savings to the design process.

We shall have more to say on this point later.

The materials interaction with design decisions is clearly an iterative one, given the complexity of the materials involvement from design development to failure analysis. Even during the earlier design stages, the materials man performs in a highly iterative mode, partly because of the interplay among the various design goals, but partly also because of uncertainties in the behavior of materials as well as in the service conditions. For example, in cases where a materials choice is difficult (e.g., high strength steel for aircraft landing gear) and where the final choice involves serious compromise with the materials engineer's sense of safety from failure, the procedures developed to assure the final reliability and cost-effectiveness of the system may require a long period of trial-and-error on the test stand.

2. Factors Affecting the Role of Materials in the Design Process

The ultimate goal of the materials engineer as he functions in the design process is to assure an efficient design. Most materials problems can be "designed around", but at a cost in performance of the system, or in the size and weight of the final product. Also, in many materials-critical applications, the ultimate performance or safety of the product, e.g., an aircraft, necessitates the most efficient and sophisticated use of materials which can be

accomplished in the design. In these instances, the trade-off in reliability vs. cost or performance is a subtle and complicated compromise. However, the importance of achieving the optimum point in this balance varies drastically as one moves from a highly materials-critical industry like aerospace to, say, the housing industry.

The criticality of materials selection in the final product is one of the main factors in determining the role of the materials engineer. The criticality may arise from the economic costs of failure, or the social costs where safety is involved, in addition to performance efficiency. Achieving the optimum compromise, however, usually entails a large amount of costly testing of materials and components under simulated life conditions; how finely tuned the compromise becomes is an economic cost which, itself, has to be taken into account in the management scheme. Obviously, the value of achieving an optimum materials selection will vary widely from industry-to-industry, and product-to-product. In this milieu, the "design solution" may turn out to be the most economical route. However, in this context, "design solution" is seen to be just another word for "over design".

Clearly, the ability of the materials engineer to perform in a quantitative way during the design process is highly desirable. When the optimum economic point is being sought, as explained above, it is advantageous for the materials engineer to be able to quantify his input. In this regard,

fracture mechanics has been very important. At the present stage, the development of K_{Ic} and K_{Isc} for the more brittle range of materials has been particularly successful; however, the application of these ideas to the more ductile materials requires further development of the J_{Ic} approach and the testing procedures that go with it. Predictive capability in the presence of hostile atmospheres is also less well-advanced, especially in the presence of stress and especially of time-dependent stress, but these subjects were not specifically explored in our workshop. These technical limitations in the ability of the materials engineer to perform quantitatively emphasize the continuing need to provide the scientific understanding upon which such practical procedures can be based.

One of the most unsatisfactory aspects of the design process at this time is NDE performance and the interpretation of NDE findings in terms of materials predictability. The workshop participants confirmed the importance that ARPA has already assigned to this field. One other external element of difficulty in the predictability of material performance is the uncertain performance envelope of the overall system. Since the actual in-life characteristics of any system are subject to many uncontrollable variables, an important concern of the materials engineer is the "forgiveness" of the material when subjected to abuse.

Finally, the materials selection process is being complicated further by a new set of externalities, e.g., the

regulatory activities of EPA, OSHA, etc., as well as the problems of materials shortages, and legal liability involving materials failures.

In the light of the increasingly complex design world facing the materials engineer, it has been proposed that a relatively small number of standardized materials, which have been tested and characterized in a large variety of conditions and performance ranges, would be a desirable objective. This proposal might have some attraction for the engineer who does not have at hand a full-scale materials back-up laboratory to perform the tests required to check out a particular materials selection adequately. However, certainly the needs of high-technology industries cannot be properly served in this way. For many applications, the continuous development of improved materials is a necessary requirement to satisfy increasingly sophisticated design goals. Also, the vast number of combinations of factors such as temperature range, environmental conditions, stress regime, etc., impose a comparable responsibility on the part of the materials engineer to choose the optimum material. Nevertheless, for low-technology purposes, the practicality of a small number of well characterized materials is worth considering.

III. Institutional and Operational Methods

The personal attributes of the materials engineer who can successfully operate in the milieu described above require

considerable development and experience well beyond the academic education of the typical materials engineer. He must become highly cost conscious; he must accept the responsibility for doing much of the trade-off between the ideal material and the practical or optimum material under normal cost and time constraints; and he must acquire a sense for the design process itself. Clearly this is continuing "education" is best carried out on the job, but it also requires considerable maturation on the part of the professional engineer himself, and corresponding effort on the part of his company.

1. High Technology Practice

We found that high technology industries have evolved an elaborate apparatus for incorporating materials input into the design process on a sufficiently high level to assure the desired result. Typically, the materials group either has the responsibility to "sign off" on the final design, or there is materials representation on a final design review board of the given company; sometimes both mechanisms are used.

Another approach is the compilation of company published materials handbooks which delineate the properties and recommended uses of materials deemed most suitable for the products being manufactured by the company. The aerospace Materials Handbook prepared by Battelle is one example of this. NASA has also produced a comprehensive materials design criteria document in which a materials matrix is developed with materials given against environmental and service conditions

in terms of the specific class of use. A representative section of this document relating to the space shuttle is appended (Attachment No. 2). In some cases, it has been found desirable to establish a complete materials tracking methodology for all materials used in a given system. This amounts to a computer cross-referencing tabulation to check an entire system for the inventory and usage (how and where) of every material contained in it. (Attachment No. 3)

We have already mentioned the on-the-job training which must take place. Many companies make considerable use of special workshops, short courses, and seminars for both materials engineers involved in the design process and for other professionals on the design team. In addition to the capabilities required of the materials engineer, similar materials awareness is required of the design engineer.

Over and beyond the intra-industry needs for communication and education, a similar issue arises when it comes to transferring technologies and materials/design procedures from one industry to another. At best, this is a slow process because individuals and groups seem to learn mainly by making their own mistakes, but some efficiency can be gained from conscious attempts to spread the results of iterative experience. We believe that short courses sponsored by societies like ASM, special simplified manuals and handbooks for designers, review articles in ASTM, etc., and publications of case studies, all have their place. We do not find that any particular form of

failure control practice should or can be legislated because of the vast differences in needs from one industry to another. The economic and social (governmental regulations, consumerism, etc.) driving forces are the ones which will engender appropriate response in a particular area and industry.

One of the special problems which arises in high technology areas is the coordination of materials control throughout the whole complex system when many companies contribute to an overall mission. The smaller vendors, especially, must be carefully monitored in terms of materials selection and control by the overall project management to insure that the materials, even in conventional devices, are appropriate to the final use.

When all is said and done, and when due recognition is given to the strides made in high technology areas to incorporate materials properly into the design process, there still remain certain hard-core problems. Central to these are program pressures on the one hand, and basic technical limitations on the other. Seldom is the optimum materials selection, recognizing realistic cost and time constraints, the ideal one from the materials standpoint alone. Judgement in the face of uncertainty is the rule, and occasional mistakes must be expected. Consequently, the pressure is always on to improve judgement by a greater degree of predictability, and a more rigorous understanding of the technical factors which underly this predictability.

2. Low Technology

At the low technology end of the scale, the style is extremely different. The design process is not characterized by materials back-up, and the design engineer himself must often function as his own materials engineer. In cases where such a design engineer encounters a situation requiring materials knowledge greater than his usual level, handbooks and manuals written to his needs would be quite helpful. Several such projects are under way, and additional help along these lines would be welcome. In this regard, the materials problem is probably not unique, and some program of periodically revising concise manuals (in very inexpensive format) for design engineers should be explored.

We have already commented on the undesirability for a major missionary effort to up-grade the materials input to low technology, where decisions are largely dominated by experience based economic balance. However, for a company which feels the need to up-grade its materials involvement in design, the necessary tools should be available.

IV. Technical Issues

We mention here for completeness some of the technical issues related to the materials/design interface, although this aspect is more adequately covered in reports by J. R. Rice and A. G. Evans.

- (1) The general materials/design situation for non-

metallics, especially polymeric materials, is much worse than for metals.

(2) The present techniques for measuring J values appear to be too conservative, since the onset of plastic crack growth is less important than the fracture instability point, where applicability of the J integral has not been established.

(3) Fatigue cracks in the ductile-fracture regime cannot yet be handled by the J-integral approach and the fundamentals of fatigue growth are poorly understood.

(4) Basic questions of how to engender toughness in high strength materials remain unanswered.

(5) Fracture in the presence of aggressive environments is still a materials-limiting roadblock.

(6) There is ultimate desirability to design materials from the basic microstructure, but present knowledge only permits qualitative guidelines.

V. Highlights and Conclusions

(1) The central conclusion of the workshop is that no major institutional "canyon" exists between the design and materials communities; by and large these two communities have learned to develop a variety of institutional responses which fit the particular circumstances.

(2) Quantitative materials predictability is at the heart of a successful materials interaction at the materials/design interface in high technology situations. This principle

was most clearly established in recent years through the medium of fracture mechanics, but major challenges remain on other fronts to improve the ability of the materials engineer to perform his function. To this end, materials research is needed toward the applicability of the J-integral for quantifying fracture toughness in service simulated atmospheres, and under time varying stress. Deeper understanding of the interrelationship between J-values and microstructural parameters is urgently needed in this whole context. A special problem exists in bringing the predictability of polymers up to the level of metals as far as the materials/design interface is concerned.

(3) NDE, and especially its application to interpretation in terms of materials reliability, require special emphasis.

(4) The techniques developed in the high technology industries to cope with materials-critical applications are: a) materials sign-off, on design, b) design-review board participation, c) preparation of special documents such as materials design monographs and handbooks, and d) on-the-job education for both materials and design engineers.

(5) There was somewhat mixed opinions regarding the desirability of new efforts at education and technology transfer mechanisms for low technology industries. Desirable activities which might be explored include: a) traveling "short lecture courses" on the materials aspects of design, perhaps sponsored by technical societies like ASM; b) selected

efforts such as the Stress Corrosion Handbook being sponsored by ARPA; and c) articles and case studies in the professional journals relevant to the materials/design interface.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

Attachment No. 1

LIST OF INVITED PARTICIPANTS IN WORKSHOP
ON THE MATERIALS/DESIGN INTERFACE

Dr. Sumio Yukawa, General Electric Company
Dr. R. Reed, NBS Boulder
Dr. L. McHenry, NBS Boulder
Dr. E. T. Wessel, Westinghouse
Dr. W. J. Levedahl, NSRDC-Annapolis
Mr. W. Dukes, Bell Aerospace
Dr. R. DeWit, NBS Washington
Dr. R. E. Johnson, Johnson Space Flight Center
Mr. A. Hurlich, General Dynamics Convair
Dr. S. Wiederhorn, NBS Washington

NUMBER	REVISION LETTER				PAGE 24
	C	D	E	F	
MC999-0096					

APPENDIX I, SPACE SHUTTLE MATERIAL SELECTION LIST, METALLIC MATERIALS

1. **DESCRIPTION:** Materials used in the design and fabrication of Space Shuttle hardware shall be selected by considering the operational requirements for the particular application and the design engineering properties of the candidate materials. Shuttle authorized materials are rated based on the specific criteria defined herein and are listed on the material selection list contained in this appendix. This list serves as the basis for all metallic material selections in both the as-designed and as-built configurations. Minimum information contained on this list shall be:

1. Material Code (Buyer assigned accession number)
2. Alloy
3. Heat Treatment
4. Generic Identification and Form
5. Material Specification
6. Maximum/Minimum Operating Temperature
7. Material Ratings for:
 - Corrosion (CRR)
 - Stress Corrosion Susceptibility (SCC)
 - Fluid Systems Compatibility
 - GOX
 - LOX
 - N2O4
 - Hydrazine (N_2H_4) and Mono Methyl Hydrazine (MMH)
 - LO PRESS. H_2 (≤ 450 psi)
 - HI PRESS. H_2 (> 450 psi)

(All ratings are based upon ambient temperature service except LOX)

2. **MATERIAL RATINGS:** Alpha characters used to rate the materials are defined as:

- A. **Acceptable.** Materials with this rating are acceptable for use without reservation in the indicated hazard category.
- B. **Acceptable with Specific Controls.** Materials with this rating are acceptable for use in the indicated hazard category provided additional specific controls are imposed. An example would be the use of 2024-T81 aluminum sheet in an environment that will cause corrosion. The specific control could be a required coating of paint. The controls are an adaption of experience or the development of new controls for Shuttle applications. The seller's design engineer will obtain the specific control information from the seller's M&P Engineering function. A "B" rated material must meet all the requirements of an "A" rating after the controls are imposed.
- C. **Acceptability Must Be Demonstrated.** Materials with this rating are not acceptable for use in the indicated hazard category until their acceptability has been established. A "C" rated material must demonstrate, by test or analysis, its ability to meet all the requirements of an "A" rating for each specific design application.

**Best Available
Copy
for page 453**

NUMBER	MC999-0096	REVISION LETTER	C D	PAGE	25
--------	------------	-----------------	-----	------	----

APPENDIX I (Continued)

U. Not rated.

X. Unacceptable. Materials with this rating are not acceptable for use in the indicated hazard category.

3. RATING CRITERIA. The criteria for rating metallic materials are as follows:

Maximum Operating Temperature

INTENTIONAL BLANK

Minimum Operating Temperature

The minimum operating temperature for metals shall be above the temperature at which ductile-brittle transition occurs, or at which a significant reduction in toughness occurs.

Corrosion Resistance

A material shall be rated "A" if it meets the requirements of MSFC-SPEC-250, Class II, and does not require a coating for its intended service.

Stress Corrosion Cracking Susceptibility

A material shall be considered rated "A" if listed as resistant on MSFC Drawing No. 1CM33107 or after 30 days of alternate exposure in a 3.5 percent sodium chloride solution (10 minutes in the solution - 50 minutes in air) at ambient temperature and pressure and stressed to 75 percent of tensile yield, and there is no failure. Failure is defined as material separation or surface cracks visible under 10 X magnification. In addition, unfailed titanium specimens shall be capable of meeting the bend test requirements of unexposed material, as defined in the applicable procurement specification, with no cracks visible under 10 X magnification.

Fluid System Compatibility - GOX

Metallic materials in GOX systems shall be rated "A" if they meet the following criteria at room temperature:

- (1) corrosion rate shall not exceed 1 mil per year per surface;
- (2) the metal passes 1500 psi pneumatic impact test in GOX per MIL 8060.1. A "B" rating is assigned to metals which fail the above test, but pass at some lower pressure. Contact buyer's M&P function for "B" rated metals. For GOX applications at pressures greater than 1000 psi, contact buyer's M&P function.

NUMBER	REVISION LETTER				PAGE	26
MC999-0096	C	D				

APPENDIX I (Continued)

3. RATING CRITERIA: (Continued)

Fluid System Compatibility - LOX

Metallic materials in LOX systems shall be rated "A" if they meet the following criteria at ambient pressure:

- (1) corrosion rate shall not exceed 1 mil per year per surface;
- (2) the metal passes the impact test per MSFC-SPEC-106.

Fluid System Compatibility - N₂O₄

Metallic materials in N₂O₄ shall be rated "A" if they meet the following criteria at 360 psi and room temperature

- (1) corrosion rate shall not exceed 1 mil per year per surface;
- (2) Stress corrosion cracking does not occur in a time representative of the system operating life at a stress level of 75 percent of tensile yield for aluminum alloys and 95 percent of tensile yield for all other alloys;
- (3) fluid decomposition is neither catalyzed nor accelerated by the metal;
- (4) the metal passes the mechanical impact test and the rapid reaction test per MSC-PA-D-67-13, addendum 2A.

Fluid System Compatibility - N₂H₄ & MMH

Metallic materials in N₂H₄ and MMH shall be rated "A" if they meet the following criteria at ambient pressure and temperature:

- (1) corrosion rate shall not exceed 1 mil per year per surface;
- (2) stress corrosion cracking does not occur in a time representative of the system service life at a stress level of 75 percent of tensile yield for aluminum alloys and 95 percent of tensile yield for all other alloys;
- (3) fluid decomposition is neither catalyzed nor accelerated by the metal.

Fluid System Compatibility - Low Pressure CH₂

Metallic materials shall be rated "A" if there is no more than 20 percent reduction in ductility or strength in 450 psi CH₂ when compared to the same test while exposed to helium.

Fluid System Compatibility - High Pressure CH₂

Metallic materials shall be rated "A" if there is no more than 20 percent reduction in ductility or strength in 5000 psi CH₂ when compared to the same test while exposed to helium.

4. TESTING. Changes to Appendix I shall be based on data from standard testing procedures of MIL 8060.1 or buyer approved procedures. All material testing shall be performed by NASA; samples shall be submitted as directed by Rockwell International to forward to NASA.

NUMBER	REVISION LETTER					PAGE	27
MC999-0096		C	D				

APPENDIX I (Continued)

5. USE OF SPECIFICATIONS.

All specifications listed in Appendix I are for reference only. Use of the specifications, for procurement, neither assures qualification to the ratings noted, nor does it imply that other materials procured to the specification have the same rating.

6. LATEST ISSUE OF APPENDIX.

As test data become available, additions are made to the material selection list (Appendix I) on a periodic basis. Seller may, at his option, request the latest issue of the Appendix to assure working with the latest information.

U719-1G-207

PAGE 1
DATE 04/28/75
MATERIAL ANALYSIS TRACKING AND CONTROL SYSTEM
MATERIAL CODE DIRECTORY BY MATERIAL DESCRIPTION

ITEMS	MATERIAL CODE	MATERIAL DESCRIPTION	SPECIFICATION	MANUFACTURER	OPER TEMP	REFERENCE	PAGE-NO	C			N H L H		
								MIN	MAX	MATERIAL	-F	PROPERTY	MANUAL
10053 AL A356-0-T61 CASTINGS				MIL-A-21180	423	5 3-13.5	5213			B-A-A-A-A-A			
10217 AL 1100 ANN WIRE		AL 1100 ANN WIRE		QC-A-225/1 QQ-A-430	423		5218			B-A-A-A-A-A			
10348 AL 1100-ANN FUEL		AL 1100-ANN FCIL		MIL-A-148 MIL-P-19834 TVI	22	LCH2/HMH2 10217	320			5212			
10279 AL 1100-ANN PLATE AND SHEET			QQ-A-250/1	22 01 10217	423	M 02.10	5212			B-A-A-A-A-A			
10346 AL 1100-C TUBING				WW-1-700/1		-320				5212			
10203 AL 2024-T6 STREET ONLY		AL 2024-T6 STREET ONLY	AMS 4C29	22 GOX 10210	423	S 3.2.1	5212			B-A-A-C-A-A			
10324 AL 2024-T3/T4 FASTENERS ONLY		AL 2024-T3/T4 FASTENERS ONLY	QQ-A-268 WW-1-785	22 01 10005		320		S 3.2.3		5211		B-A-A-A-A-A	
10216 AL 2024-142-6 TUBING				WW-1-700/3	22 GOX 10040		320		S 3.2.3	5216		B-A-A-A-A-A	
10292 AL 2024-T6 FASTENER MAIL ONLY		AL 2024-T6 FASTENER MAIL ONLY	QQ-A-225/6 QQ-A-430	22 GOX 10040		250		S 3.2.3		5211		B-A-A-A-A-A	
10291 AL 2024-T6X EXTRUSION				QQ-A-200/3	22 GOX 10040		250		S 3.2.3	5211		B-A-A-A-A-A	
10210 AL 2024-T62 SH7 & PLT		AL 2024-T62 SH7 & PLT		MB0170-078 QQ-A-250/4			250		S 3.2.3	5212		B-A-A-A-A-A	
10205 AL 2024-T6X SH7PL7(1ST T662)				MB0170-078			250		S 3.2.3	5212		B-A-X-A-A-A	
10206 AL 2024-T6X SH7PL7(2ND T662)				QQ-A-250/4			250		S 3.2.3	5211		B-A-X-A-A-A	
10207 AL 2024-T6X EXTRUSIONS				QQ-A-200/3			250		S 3.2.3	5211		B-A-X-A-A-A	
10005 AL 2024-T61 FCIL				MIL-A-01596			250						
10005 AL 2024-T61 TUBE DRILLED SEAMLES				WW-1-700/3			320		S 3.2.3	5216		B-A-A-A-A-A	

ANALYSIS OF THE DUWEZ SPLAT-COOLING PROCESS

M. Tinkham

ABSTRACT

In the Duwez drop smasher, a molten metal drop is frozen very rapidly by being compressed between two pistons driven together by compressed air. The very high cooling rate obtained in this way enables creation of solids with amorphous and unstable structures.

In this note we give an elementary analysis of the cooling and viscous effects which determine the thickness of the final disc. The analysis also gives insight about the time-dependent cooling rate, which may be useful in interpreting the structures which are obtained. Our primary conclusion is that the thermal diffusivity is the key material parameter, unless thermal boundary resistance dominates the cooling; viscosity effects are expected to be unimportant unless the viscosity is several orders of magnitude larger than the usual values of about 10^{-2} poise.

ANALYSIS OF THE DUWEZ SPLAT-COOLING PROCESS

M. Tinkham

Introduction

In the Duwez splat-cooling process, a molten metal drop is frozen very rapidly by being compressed between two pistons driven together by compressed air. The very high cooling rate obtained in this way enables creation of solids with amorphous and unstable structures.

In this note we give an elementary analysis of the cooling and viscous effects which determine the thickness of the final disc. The analysis also gives insight about the time-dependent cooling rate, which may be useful in interpreting the structures which are obtained. Our primary conclusion is that the thermal diffusivity is the key material parameter; viscosity effects are expected to be unimportant unless the viscosity is several orders of magnitude larger than the usual values of about 10^{-2} poise.

The Model

We model the drop, during its flattening, as a cylinder of radius R and height h , such that at all times $\pi R^2 h = V$, the volume of the drop. Further, until complete solidification, we assume that $h = h_0 - v_0 t$, where $h_0 \approx V^{1/3}$ is the initial

height, and v_0 is the velocity of approach of the two pistons. (This assumption will be justified by calculation of the small viscous retardation of the pistons.) Given this time-dependent thickness, we find the time-dependent cooling rate. The final thickness h_f is that which has been reached when the central temperature drops from its initial value T_0 to the melting point T_m . According to our analysis, this h_f is proportional to the thermal diffusivity D .

Viscous Effects

The initial piston velocity v_0 is estimated to be 600 cm/sec. This estimate is made by taking a pressure of 20 atmospheres acting on 10 cm^2 through a distance of 2 cm, with a piston mass M of 2000 g. The corresponding initial piston kinetic energy is 4×10^8 erg. Assuming laminar flow, this energy is dissipated at the rate

$$P_d = \eta \int \left(\frac{dv_r}{dz} \right)^2 dv \quad (1)$$

where η is the coefficient of viscosity, and v_r is the local radial velocity of the spreading drop. A simple geometrical argument shows that on the average

$$\frac{dv_r}{dz} \approx \frac{4v_r}{h} \approx \frac{2r}{h^2} \frac{dh}{dt}$$

so that the power dissipated increases rapidly with decreasing h , varying as

$$P_d \propto \eta \frac{v^2}{h^5} \left(\frac{dh}{dt} \right)^2$$

This power can be equated to the product of a resistive force F_z times (dh/dt) . Taking the numerical coefficient from the exact solution of a geometrically equivalent problem (A. H. Cottrell, The Mechanical Properties of Matter, Wiley, N. Y., 1964, p. 224), we have

$$F_z = \eta \frac{3}{2\pi} \frac{v^2}{h^5} \left(\frac{dh}{dt} \right) \quad (2)$$

Equating this to $M \frac{dv}{dt}$, and integrating, we find

$$v = v_0 - \eta \frac{3V^2}{8\pi M} \left(\frac{1}{h^4} - \frac{1}{h_0^4} \right) \quad (3)$$

$$\approx v_0 - \eta \frac{3V^2}{8\pi M h^4} \quad (3a)$$

Thus, viscous drag would stop the piston at a final separation

$$h_f|_{\text{viscous}} = \left(\eta \frac{3V^2}{8\pi M v_0} \right)^{\frac{1}{4}} \quad (4)$$

With the values taken above, and the typical value $V = 4 \times 10^{-3}$ cm³ this becomes

$$h_f|_{\text{viscous}} \approx \eta^{\frac{1}{4}} \times 10^{-3} \text{ cm} \approx 3 \mu\text{m} \quad (4a)$$

where the final result is for the typical value $\eta \approx 10^{-2}$ poise. Since the viscous retardation varies so rapidly ($\sim h^{-4}$), this result indicates that viscous effects can be neglected if the actual final thickness is greater than $\sim 6 \mu\text{m}$, as in fact it is. (To be more conservative, this h_f should be compared to

the thickness of the still-liquid central layer rather than the total thickness, but these will only differ significantly when cooling effects are more important than viscous effects; hence our qualitative conclusion should remain valid, unless the motion is so violent that the laminar flow approximation gives a dissipation level which is in error by orders of magnitude.)

Cooling Rate

Presumably the surface layers of the sample (which contact the pistons) quickly freeze, leaving a liquid central layer cooling by conduction, while simultaneously being squeezed out around the periphery to give the spreading drop. Let ΔT be the temperature difference between this central layer and the piston. If we make the approximation that the central half of the volume is at ΔT , and cooled over an area πR^2 and through a distance $(h/2)$ to the piston on either side, we have

$$-\frac{CV}{2} \frac{d}{dt} (\Delta T) = 2K \frac{\pi R^2}{(h/2)} \Delta T$$

where K is the thermal conductivity and C is specific heat per unit volume. Simplifying, we obtain

$$\frac{d}{dt} (\Delta T) = -\frac{8D}{h^2} (\Delta T) \quad (5)$$

where $D = K/C$ is the thermal diffusivity. This approach ignores the time delay associated with the diffusive aspects of the heat

flow, but since it gives a time constant $\tau(h) = h^2/8D$, which also is the approximate time required for diffusion to occur to the pistons, no serious errors should result.

Equation (5) can be integrated if we put in $h = h_0 - v_0 t$, with the result that

$$\ln \frac{(\Delta T)_0}{\Delta T} = \frac{8D}{v_0} \left(\frac{1}{h} - \frac{1}{h_0} \right) \quad (6)$$

or $\Delta T = (\Delta T)_0 e^{-\frac{8D}{v_0} \left(\frac{1}{h} - \frac{1}{h_0} \right)}$ $(h = h_0 - v_0 t)$

$$= (\Delta T)_0 e^{-\frac{t}{\tau_0 (1-t/\tau_0)}} \quad (7)$$

where $t_0 = h_0/v_0 \approx 3 \times 10^{-4}$ sec.

is the time required to flatten the drop, and

$$\tau_0 = h_0^2/8D$$

is the time for heat to diffuse out of the unflattened drop.

Taking $D \approx 0.5 \text{ cm}^2/\text{sec}$ for molten copper (from $C = 4\text{J}/\text{cm}^3 \cdot {}^\circ\text{K}$, and $K = 2\text{W}/\text{cm} \cdot {}^\circ\text{K}$) and $D \approx 0.05 \text{ cm}^2/\text{sec}$ for molten alloys, $\tau_0 \approx 10^{-2}$ sec for Cu and $\approx 10^{-1}$ sec for alloys. Thus, $\tau_0 \gg t_0$ in any case.

The final thickness h_f is found from (6) by putting $\Delta T = (\Delta T)_m$, the value at the melting point. Since $h_f \ll h_0$, we can simplify and write

$$h_f|_{\text{cooling}} = \frac{8D}{v_0} \frac{1}{\ln[(\Delta T)_0 / (\Delta T)_m]} \quad (8)$$

With $v_0 \approx 600 \text{ cm/sec}$, and $\Delta T_0 \approx 10 \text{ percent greater than } \Delta T_m$, this becomes

$$h_f|_{\text{cooling}} \approx 0.1D \text{ (cm)} \approx \begin{cases} 500 \mu\text{m for Cu} \\ 50 \mu\text{m for alloys} \end{cases} \quad (8a)$$

Considering the crudeness of our approach, these results are in reasonable agreement with values found in practice. Moreover, they are large enough to justify our neglect of viscous drag, at least for $\eta < 10^{+2}$, which would be a very high value for a liquid.

Although the drop in temperature occurs over the whole time of contact, which is only a little less than $t_0 = h_0/v_0 \approx 3 \times 10^{-4} \text{ sec}$ before freezing, the cooling rate increases rapidly as the end point is approached, varying as h^{-2} as shown by (5). The initial cooling rate (at $t = 0$) is

$$\frac{d}{dt} \Delta T|_0 = - \left(\frac{8D}{h_0^2} \right) \Delta T \quad (9)$$

which might vary from $\approx 10^5 \text{ deg/sec}$ for Cu to 10^4 deg/sec for an alloy. But at the end point, h_f , we have

$$\frac{d}{dt} (\Delta T)|_f = - \frac{8D}{h_f^2} (\Delta T) \approx - \frac{v_0^2}{8D} \ln^2 \frac{(\Delta T)_0}{(\Delta T)_m} \Delta T \approx \frac{500}{D} \Delta T \quad (10)$$

which might range from 10^6 deg/sec for Cu to 10^7 deg/sec for an alloy. It is interesting to see that, whereas the Cu cooling rate only increases by an order of magnitude during the process,

the alloy cooling rate increases by three orders of magnitude, becoming faster than the copper because the alloy drop can be flattened to a thinner shape before the center freezes. Curves contrasting the cooling rates for copper and for a typical alloy are shown in Fig. 1.

Some Necessary Refinements

The extremely simplified theory given above can not be expected to be quantitative because of the gross approximations, but it can be made somewhat more realistic by two simple extensions. For one thing, it is important to take account of the release of the heat of fusion as the drop freezes. Very roughly, this can be incorporated by replacing the specific heat C by $C^* = C + BL/(T_0 - T_m)$, where L is the latent heat and B is a numerical factor of order unity. Since L/C is of the order of 250-500°K for typical metals, C^*/C may be as large as 4 for typical superheats of ~ 100 °K. If we define an effective diffusivity by

$$D^* = \frac{K}{C^*} = \frac{K}{C + BL/(T_0 - T_m)} \quad (11)$$

We see that the predicted thickness from (8) would be reduced by a similarly large factor.

Another major refinement needed is to take account of the thermal resistance of the interface between drop and piston, and also within the piston itself. This can be done very roughly by replacing the cooling distance $h/2$ in (5) by $(h/2)+d$,

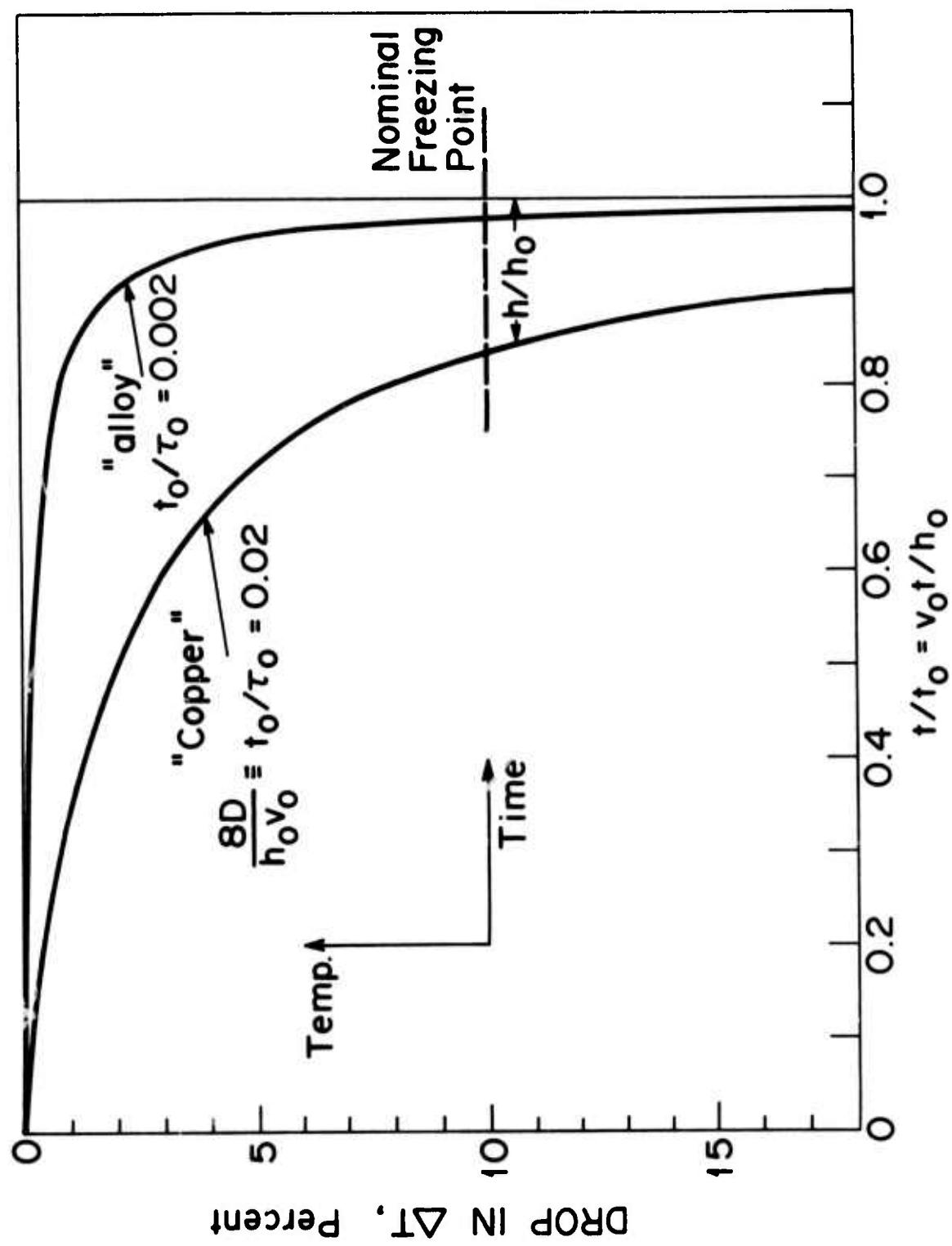


Figure 1.

where d/K equals this extra thermal resistance per unit area. If the heat transfer coefficient at the interface is α (watts/cm² °K), the interface contributes K/α to d . The thermal resistance in the piston will be approximately that of a layer of thickness equal to the thermal diffusion length after contact with the drop. Since this d is most important near the end point of the flattening, since h is then small, it is reasonable to evaluate the diffusion length at $t \approx h_0/v_0$. Thus, we have

$$\frac{d}{K} \approx \frac{1}{\alpha} + 2\sqrt{\frac{D_p h_0/v_0}{K_p}} \quad (12)$$

where K_p and D_p refer to the piston material.

Given the extra thermal resistance described by d , the differential equation (5a) becomes

$$\frac{d}{dt}(\Delta T) = -\frac{8D^*(\Delta T)}{h(h+2d)} \quad (13)$$

This also can be integrated, with the result

$$\ln \frac{\Delta T}{(\Delta T)_0} = \frac{4D^*}{v_0 d} \ln \frac{1+2d/h_0}{1+2d/h} \quad (14)$$

or
$$\Delta T = (\Delta T)_0 \left(\frac{1+2d/h_0}{1+2d/h} \right)^{4D^*/v_0 d} \quad (h = h_0 - v_0 t) \quad (14a)$$

If one is solving for h_f by setting $\Delta T = (\Delta T)_m$, then we have after some arrangement,

$$h_f = \frac{2d}{(1+2d/h_0) \exp(2d/h_{f0}) - 1} \quad (15)$$

where

$$h_{f0} = \frac{8D^*}{v_0 \ln[(\Delta T)_0 / (\Delta T_m)]} \approx \frac{8D^* (\Delta T)_0}{v_0 (T_0 - T_m)} \quad (16)$$

is the limiting value for $d = 0$. So long as $2d \ll h_0$, this expression for h_f simplifies to

$$\frac{h_f}{h_{f0}} = \frac{1}{e^X - 1} \quad (17)$$

where $X = 2d/h_{f0}$. Thus, the effect of the extra thermal resistance is to decrease h_f by a factor which decreases monotonically from unity with increasing values of $X = 2d/h_{f0}$. For the reasonable value $X = 2$, $h_f/h_{f0} = 0.31$.

In sum, since the correction for latent heat and for external thermal resistance can each decrease h_f by a factor of 2 or 3, an overall correction of up to an order of magnitude could result. These corrections would be largest for cases in which $L/C(T_0 - T_m)$ is large, or when the droplet conductivity is large, so that the thermal resistance of the interface and of the piston would be particularly important in retarding heat transfer from the droplet.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHCl5-71-C-0253 with The University of Michigan.

SUMMARY OF THE MEETING
ON FORMING SHAPES OF ADVANCED MATERIALS FROM POWDERS

E. E. Hucke

Representatives of the MRC and others from university and industry groups with research and development interest in advanced powder techniques met to discuss the state of the art, identify limiting factors, and assess the future potential of powder forming.

In his introductory remarks Professor Grant set the tone of a recurring theme; namely, that powder metallurgy with very rapidly cooled starting material offers new and exciting potential for alloy development. The more complicated of the modern alloy systems involve many purposeful alloy additions and therefore during solidification inherently produce both macroscopic and microscopic segregations. Since many of these materials can never be plastically deformed significantly, the segregations persist through heat treatment. The mean dendrite arm spacing (DAS) which characterizes the micro-segregation, can be taken as a fundamental variable setting the size scale of the structure. The DAS characteristic of ingot processes ($>10^2 \mu$) lies at one extreme while at the other is the amorphous state found in splat cooling. Between these lies a very fertile area for exploitation where the DAS $\sim 1 \mu$ to $.01 \mu$. This requires more rapid cooling rates and/or smaller

particles than currently encountered in the powder industry. Among the possible new effects are 1) existence of different minor phases, i.e., carbides and sulfides, 2) elimination of some usual phases, 3) retention of more uniform solid solutions, 4) very fine grain sizes, 5) the extension of the range of alloying elements previously prevented due to processing difficulties. As an example, 18-18 stainless with 1% S made from powder cooled at $\sim 10^2$ °C/sec. formed a very fine Cr_2S_3 phase but was cold workable, where a similar cast alloy would be very brittle. A process involving high speed inert gas atomization through a special nozzle being excited by ultrasonic vibrations was disclosed. The process was reported to have high yield of 3-5 μ powder with a present capacity of about 300 KG/hr.

Attwell Adair outlined the Air Force programs now underway toward cost reduction in various air frame and engine parts. The primary problem is one of decreasing the conversion ratio (wt. of primary material to wt. of finished part). Production of various parts from superalloys and Ti alloys has progressed so that shapes are possible that exceed the current capability for ultrasonic inspection. Large scale application of powder formed Ti parts is technically possible but awaits the availability of cheaper powder of the quality now available at $\sim \$30/\#$. Programs with Battelle utilizing their spinning process for making fine fibers and powders were described.

H. Lipsitt described Air Force efforts to meet 1980 high temperature property targets. The approach is through powder forming of Ti₃Al and TiAl compounds. With various combinations of extrusion and isothermal forging many of the program goals are at hand. Current problems lie in gaining any more than about 2% elongation at room temperature. It was, however, pointed out that some of the best Ni base alloys are now no better than 4-6% elongation.

O. Sherby discussed some interesting preliminary research on superplastic hot pressing of Fe-C alloys (~2.6%C). Particles of 25 μ containing a structure of about 2 μ were made superplastic and could easily be fully densified at temperatures far lower than usual sintering temperatures. Controlled heat treatment including malleabilization could yield some unique new Fe-C alloys.

Tom Miles described current powder surface characteristics, powder handling and superplastic hot isostatic pressing (HIP) of powder to shapes. Significant advantages were realized from continuous vacuum outgassing, assisted by an electric field. Ordinary powders were rendered superplastic by rolling the powder particles 20-40% and subsequently heat treating for a very fine structure within each particle. The need for careful control of TiC and other reactions at the particle boundaries was emphasized.

A. Cox discussed the status of powder formed parts at Pratt-Whitney, Florida. Nine different parts are in the F-100

engine and have successful service records of many thousands of hours. They have met and solved the gas contamination problem associated with current powders. This problem is less severe in the Pratt-Whitney approach than in HIP since they extrude with large reduction ratios in order to consolidate and to obtain the necessary fine structure for superplastic hot isothermal forming. This process (gatorizing) using inert atmosphere and TZM dies, gives good definition and tolerance for many complicated shapes, including discs with integral blades, blades with holes, and metal ceramic combinations.

After gatorizing, gradient annealing produced [110] oriented grains comparable in size to those directionally solidified. This material showed creep properties superior to the same alloy directionally solidified. The shapes possible have in general far exceeded those allowed by NDE, however, experimental programs are underway to establish an allowable inspection technique to be used at the extruded billet stage. Cox emphasized the need for a new alloy approach to meet 1985 goals of

200°F increased blade temperature

100°F increased disc temperature

100°F increased vane temperature

50% increase in low cycle fatigue life

100% increase in corrosion resistance

Rather dramatic immediate savings can result from improvement in fatigue strength.

R. E. Allen emphasized the need to use powder forming for reduction in cost. At present alloys cost ~\$7-8/# as a rough forging but have a conversion ratio of about 8:1. The removing of this relatively large amount of material costs about \$1-2/#. Limitation on shape in the very highly stressed parts, i.e., discs are now established by NDE and stock requirements for set up in machining. A major area for application that should be pushed is the lower stressed static parts of the engine. Allen felt that powder formed materials would not provide the best answer for the very high temperature regime. The powder formed materials are known to have lower ultrasonic background noise levels. The precise reasons for this behavior are not completely explained.

B. Avitzur discussed powder consolidation and forming under conditions of combined shear and hydrostatic pressure. Optimum combined stress conditions for obtaining sound material exist for a given material. Such conditions can successfully handle many difficult to deform materials including W at room temperature. Methods of obtaining favorable conditions in a completely continuous process called "Extrolling" were described.

M. Blackburn reviewed past alloy development in the superalloy field with reference to the possibilities for powder forming. He cited the need to develop techniques for the more modest priced alloys. From microscopic and mechanical property studies he showed how the A gas problem and the surface carbon reaction problem could be minimized in HIP processed materials.

An extended group discussion failed to identify any major scientific blind spots limiting engineering development in this field. Possible exceptions to this conclusion are 1) need for a "new" NDE method to allow the full realization of the shaping potential of current processes, and 2) the need for a better understanding of the nature of absorbed and reacted films on fine metallic powders.

In all other ways the engineering progress seems to be limited only by economic factors such as powder and processing costs. It was a general conclusion that the area of greatest future potential would come from new alloy development through utilization of known ranges of alloy content previously prohibited by processing requirements.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

TOWARD A COORDINATED PROGRAM ON
POWDER-PROCESSED MATERIALS

E. E. Hucke, J. P. Hirth
M. Cohen and P. Duwez

ABSTRACT

Impressive developments are in motion on the production and consolidation of prealloyed powders, with emphasis on very small particle sizes and fast cooling rates, thereby opening up new opportunities for compositional and structural control. Because of the ultrafine grained structures which result, superplastic behavior is induced, aiding the subsequent powder-sintering and hot-forming operations and offering good prospects for significant cost reduction in producing near-final shapes. However, the nondestructive testing of such products remains a severe problem. In addition to superfine structures, gradient annealing has produced oriented large-grained structures with creep properties superior to those obtained by directional solidification.

Through these powder processing methods, there are now good prospects for obtaining (a) unique combinations of strength and fracture toughness, particularly in titanium-base and iron-base alloys; (b) new ranges of alloy compositions (otherwise difficult to cast and fabricate) for corrosion, oxidation,

and wear resistance; (c) new approaches in the processing of refractory compounds, brittle magnetic materials, dispersion-strengthened superalloys, copper-matrix superconductors, and special stainless steels for nuclear fuel cladding; (d) greater tolerance for embrittling impurities; (e) ultrafine grain size variations bridging the transition from microcrystallinity to amorphism; and (f) tailored microstructures for elucidating phenomena relating to mechanical behavior, grain boundary diffusion, grain growth, and phase transformations.

In view of the potential advances on many fronts in both materials science and materials technology, now made possible by the powder processing methodology under discussion, we believe there is strong need for a broadly based, coordinated R&D program in this field, and we have worked out the technical and administrative features of such a plan.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

TOWARD A COORDINATED PROGRAM ON
POWDER-PROCESSED MATERIALS

E. E. Hucke, J. P. Hirth
M. Cohen and P. Duwez

I. INTRODUCTION

At the July 1975 session of the ARPA Materials Council in La Jolla, study was directed to recent developments in the processing of particulate materials, especially in superalloy technology, and the following points emerged:

1. Impressive advances are being made in the production of prealloyed powders on a substantial scale under good compositional control.
2. The importance of attaining ultrafast cooling rates in the solidification of small particles is becoming well-established as a means of achieving new orders of structural uniformity and grain refinement.
3. The resultant fine grain size leads to regimes of super-plasticity which can be brought into play advantageously during subsequent sintering and hot forming.
4. The above factors tie in closely with the newer hot-forming methods, such as hot isostatic pressing (HIP) and hot isothermal forging (gatorizing), and offer good prospects for significant cost reduction in producing near-final shapes.
5. There are also new possibilities for obtaining

superior combinations of properties in such powder-processed materials because of the unusual structural modifications and control implicit in the processing steps.

In the light of these developments, we see a whole new range of opportunities for important advances along many fronts in materials science as well as in materials technology. Accordingly, the purpose of this memorandum is to suggest a coordinated research program designed to take advantage of the fertile situation at hand.

III. GENERAL PLAN

It is suggested that ARPA initiate a coordinated, multi-contract R&D program which will include two development-oriented powder processing groups (with thrusts toward process improvement, cost reduction, and specimen preparation), and several research-oriented groups (with thrusts toward critical experiments, structure/property relationships, and new materials development). In the overall arrangement, it is intended that (a) specified materials and/or specimen-stock will be furnished by the processing groups to the research groups as necessary; (b) there will be frequent interchanges of information among the participants; (c) certain powder-processed materials may be made available to research groups outside the program contracts, as new ideas are stimulated by the central theme; and (d) an advisory committee should be appointed to assist ARPA in the technical coordination of the program.

The several components of the suggested program are discussed below, along with some indication of their relative priorities; however, nothing included here is considered to be of low priority.

III. POWDER PRODUCTION AND PROCESSING

It is proposed, as a start at least, that two powder-making and two powder-consolidating methods be brought into the program:

Powder-making

Inert-gas atomization

Ultrafast-centrifugal spinning

Powder-consolidation

Hot isostatic pressing

Hot isothermal forging

Two contractors can be chosen to produce and furnish materials by the two respective routes of (a) inert-gas atomization/hot isostatic pressing and (b) by ultrafast-centrifugal spinning/hot isothermal forging. One of these groups should have interests and experience with high-temperature materials, while the other should have similar qualifications in titanium-base, iron-base, and constructional materials. Both will not only supply materials as needed by various investigators, but may also elect to work on process variables yield efficiency, product uniformity and quality, shape control, cost analysis, and some alloy development in line with other parts of the

of the total program. Special emphasis should be given to particle-size and compositional control along with the attainment of the fastest possible quenching rates. Potential contamination at all stages of the processing should also receive due attention, and be kept to a minimum.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

ON THE POWDER FORMING OF ADVANCED MATERIALS

M. Cohen*

I. INTRODUCTION

Significant progress in the processing of particulate materials has resulted from:

- a. advances in the production of prealloyed powders of reasonable purity on a substantial scale;
- b. attainment of rapid cooling rates of the particles, thereby introducing unique possibilities for structural uniformity and control;
- c. extension of small-particle properties into the superplastic range during subsequent hot forming;
- d. advent of new hot-forming methods, such as hot isostatic pressing (HIP) and hot isothermal forging (HIF or gatorizing); and
- e. new prospects for cost savings in hot forming closer to final shapes, and also for obtaining superior properties due to unusual structural modification and control.

The above factors offer new opportunities for striking advances in materials science as well as in materials processing.

*This memorandum is a tentative summary of discussions among P. Duwez, J. P. Hirth, E. E. Hucke and M. Cohen on 3 July 1975, after the sessions on "Forming Shapes of Advanced Materials From Powders," 1-2 July 1975 in La Jolla, California.

The latter aspects will not be treated in this memorandum, but it is worth stressing that the potential economies associated with near net-shape forming will depend, in large measure, on developments in nondestructive inspection. At the present time, the ultrasonic flaw-detection methods seem to require rectilinear shapes which may be several times heavier than the final, contoured pieces. Thus, most of the hot-formed part, whatever its other merits may be, has to be machined away before the final shape is attained. This technical problem deserves specific attention.

II. NEW AVENUES OF RESEARCH

It is well recognized that the dendritic-arm spacing (DAS) in cast structures, which governs the scale of the liquid-to-solid segregation patterns, can be reduced considerably by increasing the cooling rate. For aluminum-base alloys, the following range is typical (very approximately):

<u>Solidification Method</u>	<u>Cooling Rate</u>	<u>Dendritic-Arm Spacing</u>
Large ingots	$10^{-2}^{\circ}\text{C/sec}$	$5 \times 10^2 \mu$
Gas atomization	10^2	$10^2 \mu$
Liquid atomization	10^3	5μ
Metal substrate	10^5	1μ
Splat quenching	10^7	$10^{-1} \mu$

Within any of the above categories, say atomization in argon gas, the cooling rate, and hence the DAS, depends on the droplet size. This, in turn, provides a manipulable parameter

which should make it possible (together with advanced gas-jet technologies) to move the particle size range down to 1-10 μ diameter and the DAS correspondingly smaller.

There are indications that the rapidly solidified particles, when consolidated by hot forming, can be regarded as a "novel state of matter" in some respects:

a. The rapid quenching results in a minimum of segregation, and even this is subject to fast homogenization because of the extremely small DAS which determines the interdiffusion distances.

b. At the same time, the separation of undesirable phases (embrittling compounds, impurity films, etc.) can be inhibited by the fast cooling, making possible both new alloy compositions and a greater tolerance for difficult-to-remove impurities.

c. The ultrafine grain size, inherited from the atomized particle size and persisting through the hot-forming operation, leads to superplasticity which facilitates sintering of the particles to full density along with their bulk flow to the desired shape at lower-than-usual temperatures; of course, the strain rate must be adjusted to attain the superplastic condition.

d. The effectively fine-grained and relatively uniform structures that are produced in this way lead to excellent combinations of strength and fracture properties at low-to-medium temperatures, but inferior creep characteristics at high

temperatures. However, for high-temperature applications where creep resistance becomes the limiting property, the grain size can be coarsened by heat treatment at sufficiently high temperatures, and then the expected creep resistance is obtained.

Clearly, there are many variables to be explored in the methodologies described here in order to enhance the cooling rate, to reduce the particle size and DAS, and to exercise further structural and compositional control. Undoubtedly, all these paths will be pursued along with incremental changes in alloy compositions among the existing high-temperature superalloys, toward which much of the effort has been directed so far. But what about more fundamental approaches and other opportunities for quite different applications? Here are some suggestions:

A. Phase Relationships in Alloy Systems

A number of important alloy systems should now be "revisited", particularly those which have remained in doubt because of preparation difficulties, inadequate impurity control, and undue specimen brittleness. The available powder-forming techniques should now make alloy systems accessible which could lead to sharply superior materials in corrosion and oxidation resistance. Some good possibilities for such fundamental approaches are: Fe-Al, Fe-Cr, Fe-Si, Fe-Al-Cr-Si, and Fe-Al-Cr-Si-C. (Note: Ni- and Co- base alloys are not mentioned here because they will undoubtedly be studied in various superalloy improvement programs.)

Corresponding examples in Ti-base systems are the β -phase body-centered cubic alloys, which have potential for unusual combinations of strength and fracture toughness in view of the ultrafine grain size and the capacity for martensitic transformations. The latter look especially interesting because of extended interstitial-strengthening limits and the operation of deformation-induced martensitic transformations, analogous to TRIP (transformation-induced plasticity) steels. The β -phase titanium alloys also exhibit interesting soft-mode and lattice-softening effects on cooling (prior to the martensitic transformation) which might show up exceptionally clearly with the freedom from microsegregation attainable by the powder-processing methods under discussion.

B. Some Studies of Fundamental Phenomena

Transition from Ultrafine Grain Size to Microcrystallinity and Amorphism

With continuing progress toward faster cooling rates and finer particle sizes, it is conceivable that the range from ultrafine grain size through microcrystallinity to the amorphous state can be bridged in a rather continuous way, particularly if metal/metalloid systems are chosen. The measurement and understanding of physical, chemical and mechanical changes in properties over this spectrum of solid-states would seem like a very fertile field.

Grain Growth

The availability of ultrafine grain sizes in a well-

homogenized condition with very clean grain boundaries offers a unique opportunity to investigate the mechanism and kinetics of grain growth, in which the basic interfacial-energy driving forces and atomic mobilities can dominate the grain-growth phenomena. Such studies will provide the foundation for subsequently elucidating the role of impurities and excess phases on grain boundary drag and pinning. Directional grain growth can also be studied under these conditions.

Ultrasonic Spectroscopy

It has been observed that the powder-processed bodies in question are notably more "transparent" (less background scattering) to ultrasonic waves than are the regularly cast or forged products, notwithstanding the much larger grain-boundary area in the ultrafine-grained material. Therefore, one may now wonder whether the latter grain boundaries are "transparent" because they are comparatively free of impurity concentrations, or whether there is a frequency dependence of the ultrasonic scattering. Either effect warrants study as a function of grain size and frequency in considerable detail for both theoretical and test-development reasons.

Phase Transformations

It has been reported that, in the consolidation of rapidly cooled powders, various phases are retained in solid solution which would otherwise have precipitated. This unusual condition persists despite long-time treatments at high temperatures, and raises a number of basic questions for research:

Can the equilibrium-solubility relationships be significantly changed by the powder processing, or is the effect due to drastically slower kinetics? If the former, what new thermodynamic parameters have entered the picture, and if the latter, what is inhibiting the nucleation event despite the presence of numerous grain boundaries which should be furnishing myriads of preferred nucleation sites? These very questions could lead to a new surge of theoretical and experimental activity on solid-state phase transformations.

Stress-Corrosion Cracking and Corrosion Fatigue

There is some indication in the literature that embrittlement effects, including stress-corrosion phenomena, may be related to impurity segregations at grain boundaries. Such segregations can be greatly reduced by the simultaneous advent of small particle size, fast quenching to reduce the dendritic-arm spacing, and high uniformity. It seems compelling, then, to examine stress-corrosion cracking and corrosion fatigue in carefully controlled powder-processed specimens. This approach may be one of the few remaining hopes for a substantive improvement in resistance to fracture under corrosive conditions. At the same time, some fundamental aspects of the chemical/mechanical interactions may be revealed.

C. Alloy Development

Ultrahigh-Strength Steels

The strengthening of medium alloy steels by grain refinement, superimposed on the regular martensitic strengthen-

ing, is advantageous because there is relatively little or no penalty in fracture toughness. The powder-processing route makes it worth returning to the strength versus toughness problem in ultrahigh-strength steels. Low-cycle fatigue tests should also be included in this program, particularly if the corrosion-fatigue studies mentioned above give promising results.

Corrosion and Oxidation Resistant Alloys

Obviously, with the accessibility of new compositions containing optimal concentrations of elements such as aluminum, chromium, and silicon, there will be expanded opportunities for achieving greater corrosion and oxidation resistance, concomitant with reasonable formability and more tractable mechanical properties.

Wear Resistant Alloys

The available alloys for maximum resistance to wear are usually cast to shape or are deposited as coatings because they are extremely brittle. Powder processing, for the reasons already emphasized, should add new dimensions for further alloy development and utilization in the direction of wear, abrasion and erosion resistance.

Magnets and Superconductors

Wherever the processing or application of magnetic or superconducting materials are limited by brittleness, uniformity, or particle size, it is now evident that powder forming methods of the type under consideration should receive due attention.

Refractory Materials

For similar reasons, the subject powder forming and consolidation methods should be systematically applied to the difficult-to-sinter materials, such as aluminides (e.g., Ti_3Al , $TiAl$, $NiAl$, etc.), silicides, diborides, beryllides, carbides, and nitrides. In this part of the program, molten liquid sintering agents should be invoked, where necessary or feasible.

Nuclear Fuel Cladding

The void formation and swelling characteristics of powder processed cladding alloys should be examined under radiation conditions. The results may prove to be unique.

D. Composition Control During Powder Making

An intriguing possibility is that the atomized particles can be treated in a fluidized bed with chlorine-containing argon to remove oxide films; or an aggressive atmosphere might be used to reduce the size of the particles; or chemical vapor deposition could be applied to coat the particles, either for alloying purposes or for subsequent liquid-assisted sintering. The liquid phase in the latter case could be adjusted to disappear, after serving its function, by suitable diffusion into the consolidated grains.

III. INSTITUTIONAL ARRANGEMENTS

In view of the multifaceted R&D opportunities which stem from the potential availability of well-controlled powder processed materials, consideration might be given to a coordin-

ated program in which a facility would be commissioned to furnish the necessary materials and/or specimens (in addition to carrying out its own process and alloy development) to university groups for the more basic aspects of the overall thrust. This arrangement would foster (a) a general integration of many scientific initiatives, (b) a further development of the powder technologies involved, and (c) a new mixing of the disciplines in materials science and engineering.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

QUANTITATIVE NONDESTRUCTIVE EVALUATION

ARPA/AFML Program Review

M. Cohen and R. M. Thomson

I. Introduction

An intensive three-day meeting was held at the Rockwell International Science Center on 15-17 July 1975 to review the ARPA/AFML "Interdisciplinary Program for Quantitative Flaw Definition." The scope of the meeting is indicated by a list of presented papers; see Attachment No. 1. Those papers stemming from the ARPA/AFML Program itself are designated with asterisks. The attendance was approximately 150 and the corresponding range of affiliations is shown by Attachment No. 2.

II. Nature of the ARPA/AFML Program

Dr. Donald O. Thompson of the Rockwell International Science Center is the Program Manager. There are three main Projects; each separated into Units, and each Unit having one or more assigned Tasks. The outline in Table A is useful for visualizing the technical dimensions of the overall Program, and for identifying the various groups who are participating in the coordinated research.

The objectives of the three Projects have been stated as follows:

TABLE A.

PROJECT I

QUANTITATIVE FLAW DEFINITION

UNIT I - Transducers

Task 1 Peizoelectric Transducers
K. M. Lakin, University of Southern California

Task 2 Optimization and Application of Electrodynamic Ultrasonic Wave Transducers
B. W. Maxfield, Cornell University

Task 3 Superconducting Magnets for Electromagnetic Transducers
R. B. Thompson, Science Center, Rockwell International

UNIT II - Data Processing

Task 1 Imaging and Processing of Angular Scattering
G. S. Kino, Stanford University

Task 2 Analog Data Processing
R. M. White, University of California-Berkeley

Task 3 Digital Techniques of Ultrasonic Flaw Characterization
B. G. W. Yee, General Dynamics

UNIT III - Defect Characterization

Task 1 Models for the Frequency Dependence of Ultrasonic Scattering from Real Flaws
L. Adler, University of Tennessee

Task 2 Defect Characterization by Spatial Distribution of Ultrasonic Scattered Energy
P. Packman, Vanderbilt University

Task 3 Experimental Studies of Ultrasonic Wave Scattering From Embedded Objects
B. R. Tittmann, Science Center, Rockwell International

Task 4 Theoretical Description of Scattering Phenomena
J. A. Krumhansl, Cornell University

TABLE A (continued)

UNIT IV - System Integration

Task 1 Flaw Characterization Systems
E. A. Kraut, Science Center, Rockwell International

UNIT V - Sample Preparation

Task 1 Sample Preparation
N. E. Paton, Science Center, Rockwell International

PROJECT II

BOND STRENGTH

UNIT I - Acoustical Interactions At Thin Interfaces

Task 1 Measurement of Adhesive Bond Strength with
Ultrasonic Wave Interactions
G. Alers, Science Center, Rockwell International

Task 2 Ultrasonic Signal Processing Methods for Adhesive
Bond Strength Measurement
J. Rose and P. A. Meyer, Drexel University

UNIT II - Nature of Bonded Interface Degradation in Composites

Task 1 Development of a Nondestructive Test for Strength
Degradation Composites
D. H. Kaelble, Science Center, Rockwell International

PROJECT III

FAILURE PREDICTION

UNIT I - Determination of Residual Stresses in Structural
Materials

Task 1 Harmonic Generation by Ultrasonic Surface Waves
for Materials Characterization
O. Buck, Science Center, Rockwell International

TABLE A (continued)

Task 2 Detection of Stress Through Magnetostrictive
Contribution to Electrodynamic Generation of
Ultrasound
R. B. Thompson, Science Center, Rockwell Inter-
national

UNIT II - Acoustic Emission

Task 1 Effect of Microstructure on the Frequency Content
of Acoustic Emissions
L. J. Graham, Science Center, Rockwell Inter-
national

Project I - Quantitative Flaw Definition

The tasks in Project I share the common goal of developing ultrasonic techniques necessary to quantitatively measure failure-related properties of flaws in structural materials. This involves the preparation of samples with known defect structures, basic studies of the interaction of ultrasound with these flaws, more applied studies directed towards specific schemes for processing the scattered ultrasonic energy to obtain defect information in practical situations, and studies of transducers which are the link through which the defect information must faithfully pass.

Project II - Bond Strength

The objective of this project is to investigate non-destructive methods of measuring the strength of an adhesive bond. The first of two approaches being used employs a detailed analysis of ultrasonic waves that have interacted with the adhesive bond-line to find parameters which reflect the quality of the joint. The other approach utilizes a variety of physical property measurements to define the mechanism of moisture degradation in fiber-reinforced composites so that at least one of the property measurements can be used as the basis for a reliable nondestructive test method to detect moisture degradation.

Project III - Failure Prediction

The overall objective of the failure-prediction project is to investigate nondestructive methods of measuring those

physical properties that play an important role in the failure-initiation process. Two techniques for inferring the state of residual stress are being studied. The first aims at establishing the nonlinear-elastic phenomenon of ultrasonic harmonic generation as an NDE tool for determination of the internal stress acting on dislocations and of the long-range residual stresses that produce stress gradients near the surface of a part. In order to accomplish this objective, a quantitative knowledge of the effects caused by microstructures and dislocation substructures must be established and compared to the intrinsic harmonic generation process predicted by the third-order elastic constants of the basic material. The second method for measurement of residual stress is based on the principle that a residual stress at the surface of a ferromagnetic material will bias the local magnetization and this can be measured by determining the efficiency with which a noncontact electromagnetic transducer will excite a surface acoustic wave.

III. Some Observations Concerning the Program

1. The Program, as reviewed at the 15-17 July meeting, seems well conceived, and is providing a strong technical stimulus for NDE; good science and engineering are being directed to the general problem.
2. Notable advances are being made in the fundamentals and computer-analysis of wave scattering in solids, especially

in the development of approximate methods which may apply to complex crack geometries. Excellent results are also being obtained in the area of ultrasonic imaging with multitransducer systems employing sophisticated phasing techniques.

3. Quantitative flaw detection and flaw description by ultrasonics as a reliable basis for acceptance/rejection decisions, or for failure-life calculations based on fracture mechanics, are still far off, especially with regard to cracks, but there is urgent need to pursue this line of endeavor. Very little progress appears to have been made in the ultrasonic testing of complex specimen shapes; it is necessary to deal with this problem if the full potential of NDT is to be realized. Nonmetallic materials entail even further uncertainties.

4. Metallurgical input to the ultrasonics research in the Program appears to be almost nil, except in the preparation of reference specimens with built-in defects via diffusion-bonding methods. However, some encouraging results have been obtained in the microstructural interpretation of acoustic-emission signals during plastic straining of alloys.

5. The lack of reliable flaw-detection standards is a critical roadblock at this time. NBS has an independent program in this area, and the theoretical work on ultrasonic wave scattering supported by ARPA has the potential for putting a quantitative base under this whole standards effort. The NBS and ARPA approaches should, therefore, be closely coordinated.

As an additional suggestion, we believe that the use of ultra-fine-grained powder-processed specimens should be explored for preparing standards with compositional and microstructural uniformity. Such specimens would typically produce very low background scattering of the ultrasonic waves.

6. Current piezoelectric-transducer technology and the relevant underlying understanding of the solid-state variables at play are apparently inadequate for achieving the necessary reliability and reproducibility for fine-scale flaw detection. This materials aspect of the transducer problem in NDE deserves more attention. At the same time, the noncontact advantages of electromagnetic transducers should not be overlooked, particularly inasmuch as the locally-induced eddy currents can yield structural and substructural information about the material being tested, entirely aside from flaw detection.

7. It has been reported that the Soviet Union is placing much emphasis on NTE, and that an engineering curriculum in this field is now in operation. The details of that educational approach should be examined. Indeed, eventually a technical exchange with the Russians on the subject of NDE might prove quite fruitful. This could be arranged through the National Academy of Sciences.

8. Many key areas in the ARPA/AFML NDE Program lie directly within the expertise of the ARPA Materials Research

Council and could, in fact, touch the interests of nearly every member of the Council. Accordingly, an appropriate inquiry into the critical areas of NDE should be planned for the Council's agenda next summer.

Acknowledgement

This review was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHCL5-71-C-0253 with The University of Michigan.

Attachment No. 1

ARPA/AFML REVIEW
OF
QUANTITATIVE NDE

Science Center, Rockwell International
Thousand Oaks, California

TUESDAY, JULY 15

SESSION I - INTRODUCTORY

8:45	WELCOME	D. O. Thompson, Chairman
8:50	COMMENTS FROM ARPA	T. L. Loucks, Science Center
9:00	PROGRAM DESCRIPTION & PURPOSE	E. Van Reuth, ARPA
9:10	POTENTIAL COST SAVINGS THRU ADVANCED NDE	D. O. Thompson, Science Center
9:50	DIAGNOSTIC AIDS FOR MILITARY SYSTEMS	M. J. Buckley, AFML
10:30	<u>COFFEE BREAK</u>	W. Whelan, Rand Corp.

SESSION II - ULTRASONIC SCATTERING - 1

10:45 *	QUANTITATIVE ULTRASONICS - OVERVIEW	R. B. Thompson, Science Center
11:00 *	ANALYSIS OF ULTRASONIC SCATTERING FROM SIMPLY SHAPED OBJECTS	E. R. Cohen, Science Center
11:30 *	BASIC THEORY OF ULTRASONIC SCATTERING BY DEFECTS: NUMERICAL STUDIES & FEATURES FOR EXPERIMENTAL APPLICATION	J. A. Krumhansl, Cornell Univ.

12:00 LEAKING INTERFACIAL WAVES

D. A. Lee, AFML

12:30 LUNCH

SESSION III - ULTRASONIC SCATTERING - 2

1:30 *	ULTRASONIC SAMPLES USING DIFFUSION BONDING TECHNIQUES	N. Paton, Science Center
2:00	ULTRASONIC REFERENCE BLOCKS & FATIGUE CRACKS	D. Eitzen, NBS
2:30 *	INDICIA OF SPHEROIDS & FLAT BOTTOM HOLES	P. Packman, Vanderbilt Univ.
3:00	SCATTERING OF ULTRASONIC PULSES FROM CYLINDRICAL INCLUSIONS IN ELASTIC SOLIDS	W. Sachse, Cornell Univ.

*Sponsored by ARPA/AFML Program

Attachment No. 1 (continued)

SESSION III - ULTRASONIC SCATTERING (CONTINUED)

3:30 COFFEE BREAK

3:45 *DIGITAL MEASUREMENTS OF SCATTERING FROM SPHEROIDS & FLAT BOTTOM HOLES

B. G. W. Yee, Gen. Dynamics
Ft. Worth

4:15 *COMPARISON OF THEORY & EXPERIMENT FOR ULTRASONIC SCATTERING FROM SPHERICAL & FLAT BOTTOM CAVITIES

B. R. Tittmann, Science Center

4:45 *ANGULAR DEPENDENCE OF ULTRASONIC WAVES SCATTERED FROM FLAT BOTTOM HOLES

L. Adler, Univ. of Tenn.

5:15 *REVIEW OF SCATTERING THEORIES FOR FLAT BOTTOM HOLES

E. Kraut, Science Center

6:30 ADJOURN

WEDNESDAY, JULY 16

SESSION IV - SIGNAL PROCESSING

8:45 RECENT ADVANCES IN ULTRASONIC IMAGING

S. Crist, Chairman

9:30 *APPLICATIONS OF ELECTRONICALLY SCANNED ACOUSTIC IMAGING TECHNIQUES TO NDE

R. C. Addison, American Optical

10:00 *SURFACE ACOUSTIC WAVE FILTERS FOR REAL TIME PROCESSING OF ULTRASONIC SIGNALS

G. S. Kino, Stanford Univ.

10:30 COFFEE BREAK

R. M. White, Univ. of Calif-Berkeley

10:45 RANDOM NOISE SIGNAL PROCESSING

V. Newhouse, Purdue Univ.

11:15 ADAPTIVE LEARNING APPROACH TO DEFECT CHARACTERIZATION

A. Mucciardi, Adaptronics, Inc.

11:45 INTERACTIVE PATTERN ANALYSIS & RECOGNITION

D. H. Foley, Pattern Analysis & Recognition Corp.

12:30 LUNCH

SESSION V - TRANSDUCERS

H. Bertoni, Chairman

1:30 *OPTIMIZATION & APPLICATION OF ELECTRO-DYNAMIC ULTRASONIC WAVE TRANSDUCERS

B. W. Maxfield, Cornell Univ.

2:00 ACOUSTIC SURFACE WAVE GENERATION WITH ELECTROMAGNETIC TRANSDUCERS

R. L. Thomas, Wayne State Univ.

*Sponsored by ARPA/AFML Program

Attachment No. 1 (continued)

2:30	TRANSDUCERS APPLIED TO MEASUREMENTS OF VELOCITY DISPERSION	H. Frost, AFCRL
3:00	NARROW BAND GAP SEMI-CONDUCTORS AS ACOUSTIC PHONON TRANSDUCERS	R. Francis, Univ. of Ill.- Chicago
3:30	<u>COFFEE BREAK</u>	
3:45*	PIEZOELECTRIC TRANSDUCERS FOR QUANTITATIVE NDE	K. Lakin, Univ. of So. Calif.
4:15	DEVELOPMENT OF A FIELD INSPECTION SCANNER FOR DETECTION OF CRACKS UNDER FASTENERS	W. Woodmansee, Boeing Co-Seattle
4:45	PANEL DISCUSSION	T. Cooper, Moderator
	TOPIC: "How do you foresee the requirements for ultrasonic standards changing as NDE evolves from a defect detection mode to a defect characterization mode?"	
6:30	<u>ADJOURN</u>	
7:30	NO HOST COCKTAIL HOUR AND DINNER	WEST LAKE INN - WINE CELLAR ROOM
	GUEST SPEAKER - Dr. C. Z. Serpan Nuclear Regulatory Commission	
	"CURRENT NDE PROBLEMS IN LIGHT WATER POWER REACTORS"	

THURSDAY, JULY 17

SESSION VI - ADHESIVES AND COMPOSITES	L. Lackman, Chairman
8:45 ADHESIVE INTERFACES AND FAILURE RELATED PHENOMENA - OVERVIEW	G. Alers, Science Center
9:00 FRACTURE MECHANICS OF BONDED STRUCTURES	N. Tupper, AFML
9:30 PRIMARY ADHESIVE BONDING IN ADVANCED METALLIC STRUCTURES	William L. Shelton, AFML
10:15 <u>COFFEE BREAK</u>	
10:30* DETECTION OF MOISTURE DEGRADATION IN COMPOSITES	D. Kaelble, Science Center

*Sponsored by ARPA/AFML Program

Attachment No. 1 (continued)

10:55 *INTERFACIAL STRUCTURE AND STRENGTH OF ADHESIVE BONDS T. Smith, Science Center

11:20 *ULTRASONIC WAVE INTERACTIONS WITH INTERFACES G. Alers, Science Center

11:45 *ATTENUATION INFLUENCES IN ADHESIVE BOND MODELING J. Rose, Drexel University

12:10 ULTRASONIC FREQUENCY CHARACTERIZATION OF ADHESIVE TO METAL INTERFACES J. Seydel, Univ. of Mich.

12:35 LUNCH

SESSION VII - EMISSIONS FOR NDE APPLICATIONS J. Tien, Chairman

1:30 *SOURCES OF ACOUSTIC EMISSION IN ALUMINUM ALLOYS L. Graham, Science Center

1:55 ACOUSTIC EMISSION FROM PLASTIC DEFORMATION S. Carpenter, Univ. of Denver

2:20 THEORETICAL ANALYSIS OF ACOUSTIC EMISSION SPECTRA W. Pardee, Science Center

2:45 FATIGUE ENHANCED PHOTOEMISSION INVESTIGATIONS OF ALUMINUM O. Buck, Science Center

3:10 CURRENT STATUS OF NDT WITH POSITRON ANNIHILATION F. Alex, Univ. of Utah

3:35 COFFEE BREAK

SESSION VIII - RESIDUAL STRESSES W. Walker, Chairman

3:45 MEASUREMENT OF SURFACE STRAINS BY SURFACE REFLECTION & RAMAN SCATTERING S. Ushioda, Univ. of Calif.-Irvine

4:10 *MEASUREMENT OF FLOW STRESS RELATED PHENOMENON BY NON-LINEAR ACOUSTICS O. Buck, Science Center

4:35 RESIDUAL STRESS MEASUREMENT FROM SURFACE WAVE VELOCITY DISPERSION T. Szabo, AFCRL

5:00 DEDUCING SUBSURFACE PROPERTY GRADIENTS FROM SURFACE WAVE DISPERSION DATA J. M. Richardson, Science Center

*Sponsored by ARPA/AFML Program

Attachment No. 1 (continued)

5:25	MEASUREMENT OF RESIDUAL STRESS USING MAGNETIC BARKHAUSEN EFFECT	G. Matzkanin, SWRI
5:50	*RESIDUAL STRESS BY SURFACE ACOUSTIC WAVE TRANSDUCTION	R. B. Thompson, Science Center
6:15	CONCLUDING REMARKS & ADJOURNMENT	D. O. Thompson, Science Center

*Sponsored by ARPA/AFML Program

Attachment No. 2

ATTENDEES

(As of July 7, 1975)

Mr. R. Addison
American Optical
Box 1, Research Div.
Building 7
So. Bridge, Mass. 01550

Prof. Laszlo Adler
Dept. of Physics
University of Tennessee
Knoxville, Tenn. 37916

Dr. George Alers
Science Center
1049 Camino Dos Rios
Thousand Oaks, Calif. 91360

Dr. Franklin Alex
Chief, Ind. Products Div.
Ogden ALC, USAF
Hill AFB
Clearfield, Utah

Dr. Bill Andre
Headquarters U.S. Army
Air Mobility Lab.
AIMES Research Center
Moffett Field, Calif. 94035

Mr. G. Andrews
Northrop Corp.
3901 W. Broadway
Hawthorne, Calif. 90250

Dr. Al Bahr
Stanford Research Institute
333 Ravenswood Avenue
Menlo Park, Calif. 94025

Dr. Wayne Baker
RDT Site Rep.
c/o R. O. Williams
Atomsics Intl.
8900 DeSoto Avenue
Canoga Park, Calif. 91360

Mr. Charles Bates
Brunswick Corporation
4300 Industrial Avenue
Lincoln, Neb. 68504

Dr. A. Beattie
Sandia Labs.
Org. 5143, Box 5800
Albuquerque, NM 87115

Mr. Harold Berger
Institute for Matls. Research
Reactor A-106
National Bureau of Standards
Washington, D.C. 20234

Dr. Henry Bertoni
E/EEP Dept.
Polytechnical Inst. of New York
333 Jay Street
Brooklyn, NY

Dr. Craig Biddle
Senior NDT Engr.
Pratt/Whitney Aircraft
FRDC, PO Box 2691
West Palm Beach, Fla. 33402

Mr. Bernie Boisvert
USAF NDI Prog. Mgr's Office
SAALC/MMEWA
Kelly AFB, Tex. 78291

Dr. Paul Braunlich
Bendix Research Lab.
South Field, Mich. 48075

Mr. John Brinkman
Deputy Director, RD & E
U.S. Army Armament Command
Rock Island, Ill. 61201

Mr. Alan C. Brubaker
AF Contract Mgt. Div.
AFCMD/QAC
Kirtland AFB
Albuquerque, NM 87117

Attachment No. 2 (continued)

Dr. Otto Buck
Science Center, Rockwell Intl.
1049 Camino Dos Rios
Thousand Oaks, Calif. 91360

Dr. Mike Buckley
AFML/LLP
Wright-Patterson AFB
Dayton, Ohio 45433

Dr. Steve Carpenter
Dept. of Metallurgy & Matl. Sci.
Univ. of Denver
Denver, Colo. 80112

Mr. Ed Caustin
B-1 Div., Rockwell Intl.
5701 W. Imperial Highway
Los Angeles, Calif. 90009

Dr. Marvin Chodrow
Dept. of Applied Physics
Stanford University
Stanford, Calif. 94305

Dr. R. Clements
Northrop Corporation
3901 W. Broadway
Hawthorne, Calif.

Dr. E. R. Cohen
Science Center
1049 Camino Dos Rios
Thousand Oaks, Calif. 91360

Prof. Morris Cohen
Metallurgy & Matl. Sciences
Massachusetts Inst. of Tech.
Cambridge, Mass. 02139

Mr. Don L. Conn
Sup. Research Electronics
Armco Steel Corporation
Research Center
Middletown, Ohio 45052

Mr. Tom Cooper
AFML/LLP
Wright-Patterson AFB
Dayton, Ohio 45433

Dr. Jim Couchman
Applied Research Lab.
General Dynamics
Ft. Worth, Tex. 76101

Dr. Robert Crane
AFML
Wright-Patterson AFB
Dayton, Ohio 45433

Dr. Steve Crist
Wright State University
Office of Vice Pres. for Research
Dayton, Ohio

Mr. Lee Crockett
Space Div., Rockwell Intl.
12214 Lakewood Blvd.
Downey, Calif.

Mr. George Darcy, Jr.
Chief, Nondestructive Testing
Industrialized Appl. Branch
AMMRC
Watertown, Mass.

Mr. Gary Dau
Electric Power Research Institute
PO Box 10412
Palo Alto, Calif. 94303

Dr. Robert Davis
Head, NDE Section
NASA, Lewis Research Ctr.
2100 Brookpark Road
Cleveland, Ohio 44135

Mr. Ralph Dermott
Director, Tech. Div. & Activities
American Society for Metals
Metals Park, Oh 44073

Attachment No. 2 (continued)

Mr. James Doherty
Pratt/Whitney Aircraft
Aircraft Road
Middletown, Conn. 06457

Mr. Steve Dole
Rand Corporation
1700 Main St.
Santa Monica, Calif. 90406

Mr. Don Eitzen
National Bureau of Standards
Washington, D.C. 20234

Dr. C. Elbaum
Brown University
Dept. of Physics
Providence, R.I. 02912

Mr. George Epstein
Aerospace Corp.
Bldg. A1-4081
PO Box 92957
Los Angeles, Calif. 90009

Mr. Bob Erwin
Northrop Corp.
3901 W. Broadway
Hawthorne, Calif. 90250

Dr. L. Feinstein
Stanford Research Institute
Menlo Park, Calif. 94025

Mr. Don Foley
Pattern Analysis & Recognition
On-The-Mall Center
Rome, NY 13440

Prof. Robert Francis
Dept. of Chemistry
University of Illinois
Chicago, Ill.

Col. Frischette
AFFDL/FDA
Wright-Patterson AFB
Dayton, Ohio 45433

Dr. Harold Frost
AFCRL/LZM
Hanscom AFB, Mass. 01731

Dr. Gerry Gardner
Southwest Research Institute
8500 Culebra Road
San Antonio, Tx. 78205

Dr. John Goodman
Aeronautical Systems Div./ENF
Wright-Patterson AFB
Dayton, Ohio 45433

Dr. Lloyd Graham
Science Center
1049 Camino Dos Rios
Thousand Oaks, Calif.

Mr. C. Gravette
Office of Assoc. Dir.
National Bureau of Standards
Admin. A-1002
Arlington, Va. 20234

Mr. W. S. Gray
B-1 Div., Rockwell Intl.
5701 W. Imperial Highway
Los Angeles, Calif. 90009

Dr. R. Green, Jr.
Johns Hopkins University
Mech. & Matls. Dept.
Charles & 34th Street
Baltimore, Md. 21218

Dr. Jim Gubernatis
Dept. of Physics
Cornell University
Ithaca, NY 14850

Dr. Steve Gustafson
AFML
Wright-Patterson AFB
Dayton, Ohio 45433

Mr. Steve Hart
Naval Research Lab.
Washington, D.C. 20375

Attachment No. 2 (continued)

Dr. R. E. Herfert Senior Tech. Specialist Materials Research-Northrop Corp. 3901 W. Broadway Hawthorne, Calif. 90250	Prof. Gordon Kino Dept. of Elec. Engr. W. W. Hansen Lab. of Physics Stanford University Stanford, Calif.
Prof. L. Himmel Dept. of Chem. & Matls.Sci. Wayne State Univ. Detroit, Mich. 48202	Dr. Ed Kraut Science Center 1049 Camino Dos Rios Thousand Oaks, Calif. 91360
Dr. Vince Horvath Numer Research Lab. Bethleham Steel Corporation Bethleham, Pa. 18016	Prof. J. A. Krumhansl Dept. of Physics Cornell University Clark Hall Ithaca, NY 14850
Mr. Sushil K. Jain Avco Lycoming Division 550 So. Main Street Stratford, Conn. 06497	Dr. Les Lackman LAAD-Rockwell Intl. Corp. 5701 W. Imperial Highway Los Angeles, Calif. 90009
Mr. Bill Johnson 4950/PMEA Wright-Patterson AFB Dayton, Ohio 45433	Prof. Ken Lakin Dept. of Elec. Engr. Univ. of Southern Calif. University Park Los Angeles, Calif. 90007
Dr. Dave Kaelble Science Center 1049 Camino Dos Rios Thousand Oaks, Calif. 91360	Mr. Mike Lauriente Dept. of Transportation 400-7th Street SW Washington, D.C. 20590
Mr. C. C. Kammer Space Div.-Rockell Intl. 12214 Laekwood Blvd. Downey, Calif.	Dr. David Lee Applied Meth. Research Lab. AF Aerospace Research Lab/LB Wright-Patterson AFB Dayton, Ohio 45433
Dr. Frank Kelly Principal Scientist AFML Wright-Patterson AFB Dayton, Ohio 45433	Dr. Mel Linzer National Bureau of Standards Washington, D.C. 20234
Mr. Larry Kessler Sonoscan, Inc. 752 Foster Avenue Bensenville, Ill. 60106	Dr. T. L. Loucks Science Center 1049 Camino Dos Rios Thousand Oaks, California

Attachment No. 2 (continued)

Mr. F. N. Lundberg
Science Center
1049 Camino Dos Rios
Thousand Oaks, Calif. 91360

Mr. Bill Martin
Douglas Aircraft Co.

Dr. D. C. Martin
Battelle Columbus Alab.
Mgr., Mfg. Tech Dept.
Battelle Memorial Institute
505 King Avenue
Columbus, Ohio 43201

Dr. G. A. Matzkanin
Southwest Research Institute
P.O. Drawer 28510
Tulsa, Okl. 74151

Dr. Lou Maus
Tulsa Div.-Rockwell Intl.
P.O. Box 51309
Tulsa, Okl. 74151

Prof. Bruce Maxfield
Lab. of Atomics & Solid State
Cornell University Physics
Clark Hall
Ithaca, NY 14850

Mr. D. McBride
Northrop Corporation
3901 W. Broadway
Hawthorne, Calif. 90250

Dr. F. McCuaig
Argonne National Lab.
Metallurgical Engr.
9700 S. Cass Avenue
Argonne, Ill. 60439

Dr. R. Meister
Battelle Columbus Lab.
Mgr. Quality Assurance
505 King Avenue
Columbus, Ohio 43201

Dr. Jim Miller
43929 Halcom Avenue
Lancaster, Calif. 93534

Mr. John Moore
B-1 Division, Rockwell Intl.
5701 W. Imperial Highway
Los Angeles, Calif. 90009

Dr. Tom Moran
AFML/LLP
Wright-Patterson AFB
Dayton, Ohio 45433

Mr. Tom Morton
Lockheed-Palo Alto Research Lab.
B/204, D-5231
3251 Hanover Street
Palo Alto, Calif. 94304

Dr. J. Mosier
General Electric
Dev. Project Engr.
P.O. Box 2175
Houston, Tex. 77001

Dr. Joseph Motz
Director Applied Radiation
National Bureau of Standards
Washington, D.C. 20234

Dr. Tony Mucciardi
Adaptronics, Inc.
7700 Old Springhouse Road
McLean, Va. 22101

Dr. Arthur H. Muir
Science Center
1049 Camino Dos Rios
Thousand Oaks, Calif. 91360

Dr. Max Nadler
Tulsa Div.-Rockwell Intl.
P.O. Box 51308
Tulsa, Okl. 74151

Attachment No. 2 (continued)

Dr. W. E. Nagel
Atomics International
8900 DeSoto Avenue
Canoga Park, Calif. 91304

Prof. Vernon Newhouse
Electrical Engr. Office
Purdue University
Lafayette, Ind.

Dr. Lou Odor
Columbus Div.-Rockwell Intl.
4300 E. Fifth Avenue
Columbus, Ohio 43216

Dr. Kanji Ono
Matls. Dept. 6531, BH
UCLA
405 Hilgard Avenue
Los Angeles, Calif. 90024

Prof. Ray Orbach
Dept. of Physics
UCLA
405 Hilgard Avenue
Los Angeles, Calif. 90024

Prof. Paul Packman
Vanderbilt University
Div. of Matls. Sci. & Engr.
Box 3245, Station B
Nashville, Tenn. 37204

Dr. Rod Panos
AFML/LLP
Wright-Patterson AFB
Dayton, Ohio 45433

Dr. Yih Hsing Pao
Dept. of Theoretical & Applied
Cornell University Mechs.
Ithaca, NY 14850

Dr. Bill Pardee
Science Center
1049 Camino Dos Rios
Thousand Oaks, Calif. 91360

Dr. Neil Paton
Science Center
1049 Camino Dos Rios
Thousand Oaks, Calif. 91360

Maj. Martin Peters
SAMSO (LUXT/Maj. Peters)
L.A. Air Force Station
Los Angeles, Calif. 90009

Dr. Don Pettit
Research Scientist
Lockheed-California
Dept. 74-71, Plant 2, Bldg. 20
Burbank, Calif. 91520

Mr. Jerry Posakony
Battelle North West
P.O. Box 999
Richland, Washington 99352

Mr. Joe Provens
4950/PMEA
Wright-Patterson AFB
Dayton, Ohio 45433

Dr. Howard Riess
Dept. of Chemistry
UCLA
Los Angeles, Calif. 90024

Dr. John Richardson
Science Center
1049 Camino Dos Rios
Thousand Oaks, Calif. 91360

Mr. Jeff Rohlen
LAAD, Rockwell Intl.
5701 W. Imperial Highway
Dept. 424/404, AD36
Los Angeles, Calif. 90009

Prof. Joe Rose
Drexel University
Dept. of Mech. Engr.
Philadelphia, Pa. 19104

Attachment No. 2 (continued)

Dr. W. Sachse
Dept. of Theoretical & Applied
Cornell University Mechs.
Ithaca, NY 14853

Mr. A. E. Schlemmer
Texas Eastern Transmission Corp.
P.O. Box 1612
Shreveport, La. 71130

Dr. H. Schwartzbart
Div. Matls. Engr.
400 N. Lexington Avenue
Pittsburgh, Pa. 15208

Dr. H. Schelderup
Douglas Aircraft Co.
3855 Lakewood Blvd.
Long Beach, Calif. 90846

Mr. C. Z. Serpan
Metallurgy & Matls. Branch
Nuclear Regulatory Commission
Mail Stop G-158
Washington, D.C. 20555

Dr. J. Seydel
Thriftwood Motel
575 River Parkway
Idaho Falls, Idaho 83041

Mr. W. Shelton
Northrop Corporation
3901 W. Broadway
Hawthorne, Calif. 90250

Maj. Wilbur C. Simmons
AFOSR
1400 Wilson Blvd.
Arlington, Va. 22209

Dr. Richard W. Skulski
Rockwell Intl-Rockey Flats
Bldg. 464
Golden, Colo. 80401

Mr. Hal Smallen
Aerospace Corp.
Bldg. A1-4081
P.O. Box 92957
Los Angeles, Calif. 90009

Mr. James Smith
Industrialized Applications Branch
AMMRC
Watertown, Mass.

Dr. T. Smith
Science Center, Rockwell Intl.
1049 Camino Dos Rios
Thousand Oaks, Calif. 91360

Mr. M. Stellabotte
Code 30233
Naval Air Development Center
Warminster, Pa. 18974

Mr. W. Stuhrke
Martin-Marietta
Systems Safety Studies
Mail Point 3
P.O. Box 5837
Orlando, Fla. 32805

Mr. Warren Swanson
North American Operations
1700 E. Imperial Highway
El Segundo, Calif. 90245

Dr. Thomas Szabo
AF Cambridge Research Lab.
Hanscom Field
Bedford, Mass. 01730

Dr. F. J. Szalkowski
Science Center
1049 Camino Dos Rios
Thousand Oaks, Calif. 91360

Dr. Robert L. Thomas
Wayne State University
Dept. of Physics
Detroit, Mich. 48202

Attachment No. 2 (continued)

Dr. Robb Thomson
Institute for Matls. Research
Room B354
National Bureau of Standards
Washington, D.C. 20234

Dr. Bruce Thompson
Science Center
1049 Camino Dos Rios
Thousand Oaks, Calif. 91360

Prof. Harry F. Tiersten
Rensselaer Polytechnical Institute
Mechanical Division
Troy, NY

Dr. Bernie Tittmann
Science Center
1049 Camino Dos Rios
Thousand Oaks, Calif. 91360

Mr. Nate Tupper
AFFDL/FDA
Wright-Patterson AFB
Dayton, Ohio 45433

Dr. Ed Van Reuth
1400 Wilson Blvd.
Arlington, Va. 22209

Dr. W. B. G. Yee
Applied Research Lab.
Convair Aerospace
P.O. Box 748
Fort Worth, Tx. 76101

Mr. Bill Walker
AFOSR/N4
1400 Wilson Blvd.
Arlington, Va. 22209

Mr. Tom Waugh
Dept. of Elec. Engr.
W. W. Hansen-Lab. of Physics
Stanford University
Stanford, Calif.

Dr. J. K. White
Westinghouse Research
Beullah Road
Pittsburgh, Pa. 15235

Prof. Richard White
College of Engr.-Elec. Research Lab.
University of California
Berkeley, Calif. 94720

Mr. W. Woodmansee
The Boeing Company
B-6620 73-05
P.O. Box 3707
Seattle, Wash. 98124

Dr. Tom Wolfram
University of Missouri-Columbia
223 Physics Bldg.
Columbia, Missouri 65201

Mr. R. Wolford
Northrop Corporation
3901 W. Broadway
Hawthorne, Calif.

Dr. Donald O. Thompson
Science Center
1049 Camino Dos Rios
Thousand Oaks, Calif. 91360

NON-DESTRUCTIVE EVALUATION OF CERAMICS

G. S. Kino and A. G. Evans

ABSTRACT

This paper constitutes an attempt to determine the viability of non-destructive flaw detection in ceramic materials by ultrasonic methods. As an example, the detection of flaws ($>50\mu\text{m}$ in size) in a fine-grained fully-dense silicon nitride is considered in some detail. It is concluded that a signal (from the flaw) of detectable magnitude should reach the receiving transducer if a 200 MHz input frequency is employed. The backscattering from grain boundaries will, however, prove a major source of interference, and specialized techniques may be needed to provide a clear separation of the flaw signal from the grain backscattering.

NON-DESTRUCTIVE EVALUATION OF CERAMICS

G. S. Kino and A. G. Evans

Introduction

The structural ceramic materials developed during the last few years, specifically silicon nitride and silicon carbide, have outstanding potential for use in high temperature gas turbines, bearings, etc. One of the major limitations, however, is the inability to detect critical flaws, especially in components which are subjected to complex stress states.* Typically, all flaws larger than $\sim 50\mu\text{m}$ will need to be detected in the critical (i.e., most highly stressed) parts of the component. The most obvious way of detecting these flaws is by using acoustic wave reflection techniques. This paper constitutes an attempt to determine whether testing for flaws on such a fine scale is possible in samples of the order of 1 cm thick, using hot-pressed (fully dense) Si_3N_4 as an example. The results obtained are also relevant to testing fine grained super alloys and other metals manufactured by the latest powder techniques.

*Proof testing can be very effectively used with relatively simple stress states.

Velocity in the Medium

Fully dense silicon nitride has a Young's modulus $E = 3 \times 10^{11} \text{ N m}^{-2}$, and a density of 3000 Kg/m^{-3} . The longitudinal acoustic wave velocity is thus $v = \sqrt{c/\rho} = 10^4 \text{ ms}^{-1}$, and the acoustic impedance is $Z = vp = 3 \times 10^7$. These figures are almost identical to those for sapphire.

Frequency

If we wish to determine the presence of a $50\mu\text{m}$ diameter flaw, and estimate its size and the size of larger flaws, it is probably required that acoustic waves with a wavelength, $\lambda \approx 50\mu\text{m}$ be used. This point is discussed in some detail later. The acoustic wave frequency ($f = v/\lambda$) should thus be 200 MHz.

Beam Size

We now assume that the optimum beam condition is a cylindrical beam which suffers little diffraction spreading in its path through the ceramic. The diameter D of the beam should then satisfy $D^2/\lambda > 1 \text{ cm}$; giving $D > 0.7 \text{ mm}$. We suggest a transducer mounted on a 1 cm sapphire buffer rod, tapered at one end to give a 1 mm beam.

The Transducer

It is possible to use the well established technology developed in the communications field to make a suitable transducer for this purpose.¹

A design for such a transducer is shown in Fig. 1. The transducer itself consists of a sapphire rod on which is deposited a thin film of Au on which, in turn, ZnO is deposited by sputter etching. The ZnO is a piezoelectric material which forms the actual transducer. Finally, a thin film of Au, 1 mm in diameter, is vapor deposited on the ZnO to form the contact. The sapphire buffer rod is chosen to be several times the diameter of the acoustic beam, but the end of the rod contacting the ceramic would be chosen to be 1 mm in diameter. A calculation of the equivalent circuit of this transducer yields a capacity of 5.8 pf, or a reactance of 137 ohms in series with a radiation resistance of 43.7 ohms, convenient values for matching.

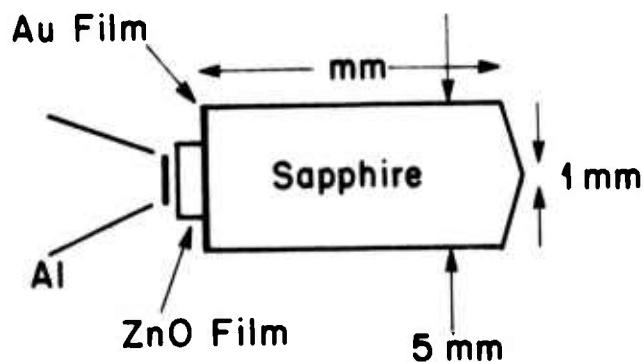


Figure 1. A schematic of a ZnO film transducer laid down on a buffer rod of sapphire.

The transducer design shown is somewhat simplified, but would be adequate for a first trial, and would probably have a conversion efficiency of less than 5 db. By using a

$\lambda/4$ Au film between the ZnO and the sapphire, good impedance matching between the ZnO films and sapphire could be obtained. Such transducers have demonstrated bandwidths of up to an octave. However, for NDE applications the transducer would be excited by a sharp pulse of the order of $2\frac{1}{2}$ nsec. Ringing must be eliminated as far as possible; so it is helpful to use a matched backing on the transducer. This would require some development, but there should be no major difficulties in making such a device.

Contacting the Ceramic

One of the difficulties of working at such a high frequency is that of making contact between the sapphire rod and the ceramic. Estimates of the pressure required to compress the asperities (Appendix) yield values that are excessively large for practical NDE testing. Therefore, some type of contacting material must be used. The choices are a liquid or a soft metal shim. Oil or water could be used, but a layer more than a few microns thick would cause a severe mismatch due to the low acoustic impedances of water or oil. A better choice would be liquid gallium, indium, or mercury. Alternatively, a thin metal shim of a soft metal like gold, tin, lead, or even copper would perform well. However, such a shim would probably have to be replaced every time the transducer is moved. NDE experiments using indium to contact the transducers have been carried out at frequencies as high as 1 GHz. Good acoustic

contact to glass was made fairly reproducibly.²

Attenuation in the Ceramic

Single crystal ceramics such as sapphire have extremely low acoustic loss and are used for delay lines at frequencies as high as 10 GHz. We might expect, therefore, that the main source of acoustic attenuation in polycrystalline ceramics would be Rayleigh scattering from the grain boundaries. We shall make a deliberately pessimistic estimate of the loss and show that it is well below tolerable levels. The primary sources of acoustic attenuation at grain boundaries are the elastic anisotropy from grain to grain in a random aggregate, and the atomic disorder at the boundary.

We regard an acoustic wave traveling through this granular medium as being a quasi plane wave with a finite cross section. The field of the plane wave is regarded as being disturbed by Rayleigh scattering in the neighborhood of a grain boundary. As we do not know the exact cause of the impedance mismatch or its value, we shall use a simplified single scatterer theory to make crude estimates of attenuation. For simplicity, we regard the crystallites as isotropic with equal shear elastic coefficients but unequal longitudinal elastic coefficients. For this case, Truell and Elbaum³ show that the Rayleigh cross section σ of an elastic sphere of radius a and acoustic impedance Z_2 in a medium of impedance Z_1 , is

$$\sigma = \frac{4}{9} (ka)^4 \left[\left(\frac{Z_2}{Z_1} \right)^2 - 1 \right]^2 \pi a^2 \quad (1)$$

where k is the propagation vector ($= 2\pi/\lambda$).

The number of scatterers per unit area is $1/\pi a^2$ and thus, the effective cross section of the scatters per unit area is

$$\sigma_{\text{tot}} = \frac{4}{9} (ka)^4 \left[\left(\frac{Z_2}{Z_1} \right)^2 - 1 \right]^2 \quad (2)$$

The quantity σ_{tot} is essentially the proportion of the incident wave intensity, I , that is scattered. The power lost for a distance of travel equal to the grain diameter ($2G$) is thus, simply, $\sigma_{\text{tot}} I$, or

$$\left(\frac{dI}{dz} \right) = \sigma_{\text{tot}} I / 2G \quad (3)$$

where z is the distance traveled by the wave. Hence the attenuation, α , of the wave per unit length is:

$$\begin{aligned} \alpha &= \frac{1}{2I} \left(\frac{dI}{dz} \right) = \frac{1}{9G} (kG)^4 \left[\left(\frac{Z_2}{Z_1} \right)^2 - 1 \right]^2 \\ &\equiv \frac{16\pi^4}{9\lambda^4} G^3 \left[\left(\frac{Z_2}{Z_1} \right)^2 - 1 \right]^2 \end{aligned} \quad (4)$$

If we now choose $1.5\mu\text{m}$ as a typical grain diameter for silicon nitride, and we estimate the average elastic anisotropy, $(c_{11})_1/c_{11})_2$, to be at most 1.5, giving $(Z_2/Z_1)^2 < 1.5$, the estimated attenuation is 0.25 db/cm, a negligible loss. However, since the loss varies as f^4 , if the frequency were raised to 400 MHz, the estimate would become 4 db/cm.

It is difficult to check these estimates because very little information on scattering loss in ceramics is available. A direct measurement of scattering loss of surface acoustic waves in PZT ceramics has been made using laser techniques, by De Vries.⁴ In this case, a prime source of loss is surface scattering, which varies as f^5 . His measurements should thus yield an attenuation far greater than the value due to grain scattering losses. The measured acoustic surface wave attenuation (at 46 MHz) in a sample of fine grain Honeywell type S PZT was .024 db per wavelength, or 4.8 db/cm. The surface wave velocity in this material is 2.3×10^5 cm/sec, so the wavelength, $\lambda = 50\mu\text{m}$, happens to be identical to the wavelength we have suggested for the Si_3N_4 . Thus, we might expect the grain scattering losses to be similar to those expected at 200 MHz in Si_3N_4 . The only difference may be the elastic anisotropy; however, even assuming that the anisotropy ratio $[(c_{11})_2 / (c_{11})_1]_{\text{average}}$, for the PZT is only 1.1, we would estimate the grain scattering loss (for a $5-7\mu\text{m}$ grain diameter) to be 4-11 db/cm. This is comparable to or larger than the experimental value.

Hence, we conclude that our method of estimation of attenuation is probably pessimistic and there should be no difficulty in operating at 200 MHz as far as acoustic losses are concerned. Furthermore, it may be quite possible to operate at 400 MHz (a $25\mu\text{m}$ wavelength).

Scattering from a Flaw

We consider, for simplicity, a spherical flaw of radius a , a distance z from the transducer, as illustrated in Figure 2.

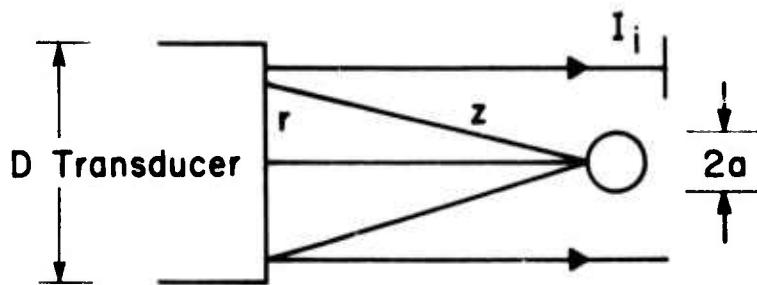


Figure 2.

If the acoustic intensity of the incident wave is I_i , the reflected signal intensity I_R at the transducer can be shown for $\lambda \gtrsim 2a$ to be⁵,

$$I_R = \Gamma^2 I_i \frac{a^2}{4z^2} \quad (5)$$

where Γ is the reflection coefficient of the flaw. The rays reflected to the transducer will have phase delays along them, determined by the ray length. Suppose we consider the signal reaching a point a distance r from the axis of the transducer. Then the excess phase delay of the ray reaching this point over that passing along the axis is, by the paraxial approximation,

$$\phi = \frac{\pi r^2}{\lambda z} \quad (6)$$

As an approximation, only those rays with $\phi < \pi/2$ will contribute to the signal exciting the transducer. A more detailed analysis involving Fresnel integrals confirms this result. Thus, the useful area A of the transducer is given by the relation

$$A = \pi r_{\max}^2 = \frac{\pi \lambda z}{2} \quad (7)$$

If the power radiated from a transducer of diameter D is $P_i = I_i \pi D^2/4$, the reflected signal received at the transducer is $P_R = AI_R$. Thus we find that

$$\frac{P_R}{P_i} = \frac{\lambda a^2}{2D^2 z} \Gamma^2 \quad . \quad (\lambda \gtrsim 2a) \quad (8a)$$

Note that for wavelengths substantially above the flaw diameter, a different relation applies (c.f. Eq. (12)*;

$$\frac{P_R}{P_i} = \frac{32\pi^5 a^6}{2D^2 \lambda^3} \left[\left(\frac{z_f}{z_m} \right)^2 - 1 \right]^2 \quad (\lambda \gg 2a) \quad (8b)$$

where z_f and z_m are the acoustic impedance of the flaw and the matrix, respectively. When working with wavelengths comparable to the flaw diameter we find from Eq. (8a) that, in the maximum reflectivity case ($\Gamma = 1$, a void) and for $D = 1\text{mm}$, $a = 25\mu\text{m}$, $\lambda = 50\mu\text{m}$, $z = 2\text{cm}$, the reflected signal is attenuated by -61 db. If we consider the total transducer and contact loss to be 15 db the return signal will be 76 db below the incident signal level. Thus, with a 1 watt input pulse the output would be

*This equation differs from the normal λ^4 relation because the frequency dependence of the useful area of the transducer has been incorporated in the present analysis.

-46 dbm. The best way to detect the short pulse required would be to use a sampling oscilloscope. Sampling oscilloscopes with a 50 ohm input impedance typically have a minimum sensitivity of 2 mv. This corresponds to an input power of -80 dbm. Thus on this basis, assuming negligible loss in the ceramic, a 50 μ m void would give a return signal 34 db above noise. A defect with less reflectivity than a void would be correspondingly harder to detect.

The return signal from the smallest flaw size of interest is seen to be weak, although detectable. This is partly because we have suggested the use of far lower peak powers than are commonly employed at low frequencies. However, the proposed transducers and the medium matching lead to a higher electrical/acoustic coupling efficiency than might be expected at low frequencies. Here, under good conditions, there is certainly the possibility of a further 10 db improvement.

In order to obtain a larger dynamic range, so that the smallest flaws are very easily detectable, further measures to improve the signal level are possible. It is certainly possible to drive the transducer at a 100 watt level, although some type of protection will then definitely be needed for the sampling oscilloscope. In this case, it is possible to use an input FET amplifier with sufficient protection, or to use diode protection. The recovery time of the protection circuit with the use of a 1 cm sapphire buffer rod need only be 3 μ sec.

A second possibility for which the signal processing system is well developed is to use an FM signal with a linear variation of frequency with time as the exciting source. The return signal is then passed into a matched dispersive filter, consisting of an acoustic surface wave delay line. By using an FM signal of the same bandwidth as that required in the pulse (200 MHz), and a time length of $T = 1\mu\text{sec}$, it is possible, because of the long time scale used, to obtain an improvement in signal to noise ratio of T/τ , where τ is the length of the equivalent pulse (5 nsec). This implies an improvement in signal to noise ratio of 200:1, i.e., 23 db, using the same 1 watt input signal. The required signal processing technology is used in radar systems and is available. In fact, much greater improvements of as much as 5000:1, or 37 db, are possible, although the delay lines required are not as easily available.

Back Scattering from Grain Boundaries

The most severe problem in making measurements of scattering from small flaws is that associated with back scattering from the grain boundaries. We shall try to estimate the magnitude of this effect for grain diameters substantially smaller than the wavelength by taking account of the scattering from individual grains, and regarding the scattering as being random in phase (assuming a random grain orientation and hence, a random impedance mismatch at grain boundaries).

For this reason, and for simplicity, we shall consider the average scattering from an individual grain to be isotropic. Thus the power, P_r , scattering into the solid angle Ω by an individual grain is:

$$P_r = I_i \sigma \Omega / 4\pi \quad (9)$$

Now referring to Fig. 2, the power received in a circular area of radius r in the transducer, whose axis passes through the grain, with the transducer a distance z from the grain, is

$$P = \frac{r^2}{4z^2} I_i \sigma \quad (10)$$

Following the argument that we have already employed for scattering from a flaw, the effective radius r_{\max} of the illuminated region on the transducer, where the ray paths have less than $\lambda/4$ difference in length, is $\lambda z/2$. Hence, the power reaching the transducer from a single grain is

$$P = \frac{\lambda}{8z} I_i \sigma \quad (11)$$

Expressing I_i in terms of the radiated power P_i , Eq. (11) becomes:

$$P = \frac{P_i \lambda \sigma}{2\pi z D^2}$$

where D is the transducer diameter. Here we have assumed that $r_{\max} < D/2$, as it would be when the beam length has been chosen to be within the Fresnel region. We are over-estimating the back scattered power received at the transducer by assuming

that the area circumscribed by a circle of radius r_{\max} lies within the transducer diameter.

If the grains have a volume $(4/3)\pi G^3$ the total power received at the transducer from a short length L of the sample, i.e., the region from $z - \frac{L}{2}$ to $z + \frac{L}{2}$, when illuminated by a beam of diameter D , is given by the relation

$$\frac{P_R}{P_i} \approx \frac{\lambda \sigma}{2z\pi D^2} \frac{3L D^2}{16G^3} \approx \frac{3\lambda L \sigma}{32\pi G^3 z} \quad (13)$$

Substituting the expression for σ from Eq. (1) we find that

$$\frac{P_R}{P_i} = \frac{2\pi^4 G^3 L}{3\lambda^3 z} \left[\left(\frac{z_2}{z_1} \right)^2 - 1 \right]^2 \quad (14)$$

We now consider using the minimum possible pulse length L (approximately equal to the wavelength λ), take $\lambda = 50\mu\text{m}$, $z_2^2/z_1^2 = 1.5$, $a = .75\mu\text{m}$, $z = 1 \text{ cm}$ (minimum value), and find that the estimated scattered power is -66 db down from the incident power. This is an extremely low figure, and is 5 db less than our estimate of the scattering from the smallest flaw which we may wish to observe.

The preliminary calculations thus tend to indicate extremely low values of backscattering from a short rf pulse. The use of a long rf pulse would lead to more pessimistic conclusions. We note from Eq. (14) that where $L = \lambda$, then $P_R/P_i \propto 1/\lambda^2$. So a reduction in frequency to 15 MHz would decrease the backscattered power by 23 db. But, at the same time, the power reflected from a $50\mu\text{m}$ flaw (which varies approximately as $1/\lambda^3$, Eq. (8b)) would be reduced by 34 db.

A reduction in frequency thus diminishes flaw detectability. Additionally, comparison of Eqs. (8a) and (8b) indicates that for wavelengths smaller than the flaw diameter the power reflected from the flaw diminishes in direct proportion to the wavelength; also, as the wavelength decreases, the attenuation increases (Eq. 4). Hence, very high frequencies are equally undesirable. We conclude, therefore, that it is usually wiser to work at a frequency whose wavelength is comparable to the flaw size.

One question that arises is the effect of multiple scattering from the grains. This is difficult to estimate. The number of scatterers which can give a second order interaction of this type is approximately the number in the beam between the entrance plane into the medium and the plane where the pulse is present. If $z = z_1$ is the entrance plane, this corresponds approximately to $(z - z_1)D^2/4G^3$ scatterers.

Now the time delay from the first scatterer at \bar{R}' to the second scatterer at \bar{R}'' and then back to the transducer must be such that if the return signal arrives at the same time as the return pulse from the plane z , then it follows that $|\bar{R}''| + |\bar{R}'| + |R'' - R'| = 2z$, as shown in Fig. 3.

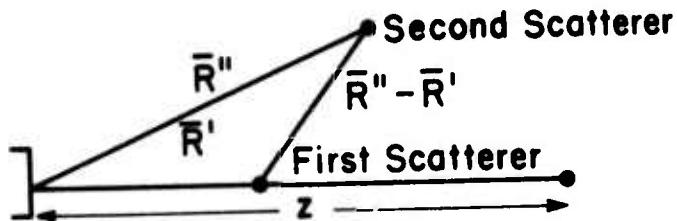


Figure 3. A schematic illustrating multiple scattering.

If we assume that the second scatterer must lie within the beam so that the transducer can respond to it, then the second scatterer must be within the region $z-L/2 < R'' < z+L/2$, i.e., within the beam pulse, for single scattering back to the transducer.

We shall estimate the effect of multiple scattering by estimating the power reaching the plane z (the location of the second set of scatterers), from the first set of scatterers. We determine the power intensity of the single scattered energy on the axis at the plane z due to all the scatterers in the beam from the entrance plane z_1 to z . If this is far less than the direct power intensity reaching this plane (without scattering), we can neglect multiple scattering effects.

We know a maximum possible value for the scattered power, the power lost by attenuation due to scattering. This implies that, with use of the pessimistic assumption, that half of the scattered power reaches the plane z (all the power scattered in the forward direction).

$$I(z)_{\text{scattered}} = \frac{1}{2} I_i \left[1 - e^{-2\alpha(z-z_1)} \right] \quad (15)$$

Taking $z-z_1 = 1$ cm, we find from the previous calculation that Eqs. (4) and (15) yield the result $I(z) = .025 I_i$. So, as an estimate, 0.025 of the incident power would be available for second scattering. Thus, even if second scatterers well outside the main beam were contributing to the return backscattered signal, the second order scattering would only be com-

parable to the first order scattering if the area of the scattering region was $(1/.025) = 40$ times that of the beam. This would be equivalent to a disc of diameter 6 mm, far too big for most of the secondarily scattered power to excite the transducer of 1 mm diameter. Thus, we can neglect multiple scattering.

Conclusions

We conclude that it should be possible to carry out acoustic reflection measurements in hot pressed silicon nitride to detect $50\mu\text{m}$ flaws. The backscattered radiation is calculated to be of the same order as the signal from the flaw, and the attenuation to be negligible at a frequency of 200 MHz.

The use of higher or lower frequencies leads to a relatively higher level of backscattered radiation, compared to the signals reflected from the flaw, i.e., optimum detectability is achieved by working with wavelengths comparable to the flaw diameters.

The calculations are based on an assumed grain size of $1.5\mu\text{m}$, a beam size of 1 mm and an average mismatch in elastic constants at a grain boundary of 1.5:1. If the grain size were larger, the grain scattering would be correspondingly more serious; both backscatter radiation and loss vary as the cube of the grain diameter.

On the other hand, our estimate of the average mismatch in impedance may well be far too high. If this is the case,

the experiments may be easier to perform than the theory would indicate.

There are possibilities for improving the signal level from a flaw and decreasing the backscattered radiation. The use of a focused beam would be of great help in this respect. A reduction in the beam diameter at the flaw by a factor of 5 (to 0.2 mm) would increase the reflected power by a factor of 25 or 14 db and decrease the scattered power by the same factor, thus improving the signal to scattered noise ratio by 28 db.

We suggest that initial experiments should be carried out to measure the acoustic attenuation and backscattered radiation. If the results are comparable to or worse than the estimates made here indicate, it may then be necessary to resort to more sophisticated techniques. In this case, the best possibility would be to use a focused beam, or to combine the use of a movable transducer with computer correlation techniques of the type suggested by Seydel. This technique is entirely equivalent to using a movable lens. In the end, it may be necessary to employ electronically focused lens techniques in order to obtain the required information in reasonable times.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

APPENDIX

The Mechanical Contact Between the Component and the Wave Guide

Consider the contact between a mechanically polished sapphire wave guide and a machined component. The extent of the mechanical contact is determined by the elastic deformation of the surface asperities on the component. For an asperity of radius, r_i , the contact radius, a_i , for a force, P_i , is

$$a_i^3 \approx \frac{P_i r_i}{E} \quad (A1)$$

The contact area A_i is thus;

$$A_i = \pi a_i^2 = \pi \left(\frac{P_i r_i}{E} \right)^{2/3} \quad (A2)$$

The total area of contact for N asperities is then;

$$A = \frac{\pi}{E^{2/3}} \sum_{i=1}^{i=N} (P_i r_i)^{2/3} \quad (A3)$$

and the areal fraction of the wave guide cross-section in contact with the component, f , is;

$$f = A/\pi \sum_{i=1}^{i=N} r_i^2 = \frac{1}{E^{2/3}} \sum_{i=1}^{i=N} \left(\frac{P_i}{r_i^2} \right)^{2/3} \quad (A4)$$

But, the total stress, σ_a , applied to the wave guide is

$$\sigma_a = \frac{1}{\pi} \sum_{i=1}^{i=N} \frac{P_i}{r_i^2} \quad (A5)$$

Combining Eqs. (A4) and (A5) yields the simple result

$$f^{3/2} = \frac{\pi \sigma a}{E} \quad (A6)$$

The contact area is thus independent of the absolute size of the asperities.

Limitations on the maximum contact are imposed in practice by localized deformation and fracture. For small asperities $\lesssim 20\mu\text{m}$, deformation occurs prior to fracture (although a more gross form of fracture - Hertzian fracture - from the periphery of contact between the wave guide and the component can intervene for badly machined surfaces). Asperity deformation is determined by the material's hardness, H , which is given by;

$$H = \frac{P_i}{\pi a_i^2} \quad (A7)$$

For structural ceramic materials H is typically $0.05E$. The maximum force that should be applied to each asperity to prevent permanent surface deformation is thus, $P_i < 0.05 \pi E a_i^2$. Thus from Eq. (A1)

$$\left(\frac{a_i}{r_i}\right)_{\max} = 0.05 \quad (A8)$$

giving;

$$f_{\max} \approx 5 \times 10^{-3} \quad (A9)$$

and, hence, from Eq. (A6)

$$(\sigma_a)_{\max} \approx 10^{-4} E \quad (A10)$$

Thus, relatively small contact areas (Eq. A9) can be achieved without surface degradation; even then the stress levels are quite large, e.g., $\sim 40 \text{ MN m}^{-2}$ for a typical structural ceramic. Good mechanical contact between the wave guide and the component does not appear, therefore, to be a practical possibility.

REFERENCES

1. T. M. Reeder and D. K. Winslow, "Characteristics of Micro-wave Acoustic Transducers for Volume Wave Excitation," IEEE Trans. MTT-17, 927-941 (1969).
2. W. Yee, "Microwave Acoustic Time-Domain Reflectrometry," E. E. Thesis, Stanford University, 1972.
3. R. Truell, C. Elbaum, and B. B. Chick, "Ultrasonic Methods in Solid State Physics," Academic Press, New York, 464 p. (1969).
4. A. J. DeVries and R. L. Miller, "Optical Measurements of Surface-Wave Scatter Losses in Piezoelectric Ceramics," Appl. Phys. Lett. 20, 210-212, 1972, and private communication.
5. R. J. Urick, "Principles of Underwater Sound for Engineers," McGraw-Hill, New York, 342 p. (1967).

COAL LIQUIFACTION AS AN INTERCALATION PROBLEM

R. A. Huggins

ABSTRACT

Some of the experimental information on the structure and the liquification of coals is reviewed. This is discussed in the light of current knowledge of the intercalation of species between the carbon sheets in graphite and other groups of materials with layer structures.

Interaction between intercalated species and the host layers, structural changes, exfoliation, and some unique properties of intercalated layer structures are also mentioned.

COAL LIQUIFACTION AS AN INTERCALATION PROBLEM

R. A. Huggins

INTRODUCTION

There are a number of technological problems that must be overcome in order for coal to become a viable source of some of the important ingredients necessary for the production of feedstocks to be used in the manufacture of petrochemical products. One of the possible approaches involves the liquification of coal, and various schemes, such as the "SRC Process", the "H-coal Process", and the "Synthoil Process", are being developed for this purpose.

It is the purpose of this note to point out several observations related to the liquification of coal that indicate that, in certain cases, it may be useful to consider analogies with what is known with regard to the intercalation of species in graphite and other layer compounds.

THE STRUCTURE OF COAL

It is well recognized that both the composition and the structure of coal vary greatly, depending upon its geological source, and several reviews of current knowledge are available.¹⁻⁶ Although there is evidence⁷ that in some coals, aliphatic groups, rather than polynuclear condensed aromatics predominate, in other cases⁸⁻¹¹ it appears that polycondensed

aromatics are more prevalent. There is also evidence¹¹ for a preferred orientation of aromatic layers parallel to the bedding plane.

Despite these apparently conflicting results, it seems reasonable to draw a number of conclusions. The ratio of aromatic to aliphatic material increases sharply with the rank of the coal. In the lower-rank coals the diameter of the aromatic lamellar is quite small (4-6 condensed rings per layer), and contains about 70% of the total carbon present. These layers are present singly or in small groups, may be somewhat puckered, and have a slight preferred orientation. There are also many anisotropic pores in the structure, which tend to be parallel to the bedding plane.

In higher ranked coals the diameter of the aromatic layers increases, as does the number of parallel layers. The structure becomes more uniform and less open, with fewer pores. In coal with a total carbon content of 94%, the structure is almost perfectly graphitic, with an interlayer spacing of 3.43\AA and a distance of 1.42\AA between carbon atoms. The degree of preferred orientation is also quite high. These changes in structure with rank are illustrated schematically in Fig. 1, taken from the paper by Hirsch.¹¹

The structural constituents which tie the groups of aromatic layers together are evidently primarily aliphatic (e.g., methylene and sulfur-sulfur) bridges. This material is quite volatile, and its amount decreases sharply with higher

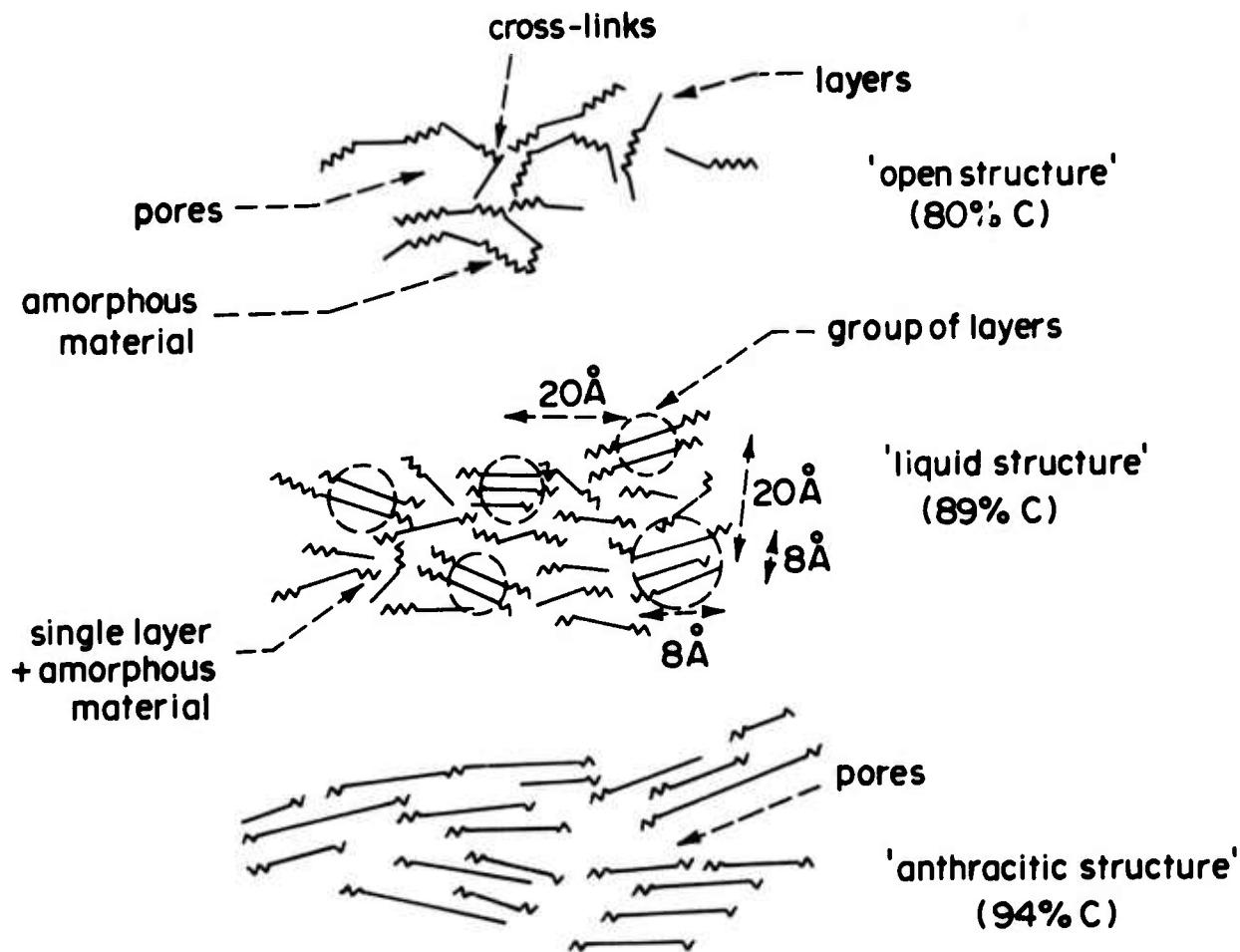


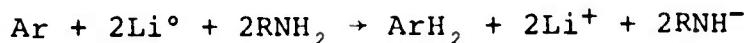
Figure 1. Changes of stacking order with rank.

rank. It also appears that the local graphitic regions are packed together randomly: Hirsch¹¹ describes this as liquid-like. Interestingly, this liquid-like arrangement seems to be most prevalent at a composition of 89% carbon. Porosity again appears in coals of higher rank (anthracites). It seems that the inter-layer forces are weakest in coals at 89% carbon. There is a minimum in both Young's modulus and the high temperature viscosity in coals of this rank, which have the best coking properties. Flow may actually take place by slipping of the layers relative to each other.

SOME COMMENTS ON THE SOLUBILIZATION OF COAL

Evidently, solvents in which coals are the most soluble are those which act as Lewis acids (electron acceptors), and also contain aromatic species. An example of a solvent that is used for this purpose is $ZnCl_2$ in an organic host, another is boron trifluoride in phenol. It is found⁴, however, that the solubility of coal in phenol-boron trifluoride varies with rank, being greater for coals of lower rank.

The solubility of coal in solvents such as pyridine can be greatly enhanced by hydrogenation of the aromatic layers.⁴ This can be performed in solutions of alkali metals in solvents like ethylenediamine. The following reaction is assumed for this reaction:



where Ar is the aromatic layer in the coal.

It has also been found¹²⁻¹⁴ that solubility can be enhanced by electrochemical hydrogenation, using electrolysis of lithium chloride in ethylenediamine. In this case, lithium ions are described⁴ as catalysts for the hydrogenation. This behavior is similar to the electrochemical reduction of aromatic hydrocarbons.¹⁵⁻¹⁷

In general, it is found that alkylation of coal greatly increases the solubility in organic solvents such as pyridine and benzene. This evidently due to the insertion alkyl groups between the aromatic layers, in some cases causing puckering. This solvation increases the ease of separation of the layers, and hence, the solubility of this part of the coal structure.

It is also found that the solubility of coal is greatly dependent upon the dielectric constant of the solvent.

ANALOGOUS INTERCALATION OF GRAPHITE AND OTHER MATERIALS WITH LAYER STRUCTURES

The phenomena that have been mentioned in the previous section seem to be quite analogous to many observations that have been made by those working on the intercalation of species between the sheets of materials with well-defined two-dimensional layer structures.

There is a rich literature on the intercalation of species between the aromatic layers in graphite and on various compounds formed in this way by graphite. In some cases species intercalated between the aromatic lamellar can be highly mobile.

There will be no attempt to review these matters here; quite thorough reviews can be found elsewhere.¹⁸⁻²²

To mention just one point, it is interesting that although the alkali metals potassium, rubidium and cesium readily enter the graphite structure, sodium and lithium do not easily do so at low temperatures. On the other hand, it has recently been shown²³ that lithium can be driven into graphite by electrolysis of molten salts, and causes exfoliation and fracture of the graphite structure. Exfoliation can also be caused to occur by a number of other means, and the term "extercalation" was introduced²²⁻²⁴ to describe compounds in which this occurs.

Analogous phenomena occur in other groups of materials with layer-type crystal structures in which the bonding within the layers is rather strong, and between layers relatively weak. Among materials of this type are some transition metal dichalcogenides, such as TiS_2 and TaS_2 . Because of an interest in their superconducting properties, and more recently, in their possible use as cathode constituents in battery systems and for catalytic purposes, this group of materials has received a considerable amount of attention. Recent reviews can be found in several places.²⁵⁻²⁷ Work in these systems has shown the great variety of species, both inorganic and organic, that can be put between the lamellae in such structures. Molecules that act as Lewis bases seem to be most amenable to intercalation, and well-defined structures have been produced in this manner²⁵ with interlamellar spacings as great as 57 \AA .

In both graphite and these other incalatable systems it has been found that intercalation can be reversible, a sequence of different species can be placed between the layers, and easily insertable "carrier species" can be used to "pry open" the structure in order to allow the entry of other species.

SUMMARY

Because of the similarity in both structural features and observed phenomena, it seems that it would be useful to further consider the possible relations between the problem of the liquification of at least the higher rank coals and current knowledge concerning both graphitic and other intercalation compounds.

ACKNOWLEDGEMENT

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

References

1. W. R. K. Wu and H. H. Storch, "Hydrogenation of Coal and Tar," *Bur. Mines Bull.* 633 (1968).
2. D. W. Van Krevelen, "Solvent Extraction of Coals," Chapt. X in Coal-Typology, Chemistry, Physics and Construction. Elsevier Publishing Co., New York (1961) pp. 177-200.
3. G. R. Hill and L. B. Lyon, "A New Chemical Structure for Coal," *Ind. Eng. Chem.* 54, 36-41 (1962).
4. I. Wender, "Characteristics of Coal Behavior in Chemical Reactions," unpublished material used in course at Penn. State Univ.
5. E. E. Donath, "Hydrogenation of Coal and Tar," in Chemistry of Coal Utilization, Suppl. Vol. (Ed. H. H. Lowry), Wiley, 1963.
6. Liquifaction and Chemical Refining of Coal, Battelle Energy Program Report, Battelle Columbus Laboratories, Columbus, Ohio (1974).
7. R. A. Friedel and J. A. Queiser, *Fuel* (London) 38, 369-79 (1959)
8. J. K. Brown, *Fuel* (London) 38, 55-63 (1959).
9. J. K. Brown, W. R. Ladner, and N. Sheppard, *Fuel* (London) 39, 79-86, 87-95 (1960).
10. P. H. Given, *Fuel* (London) 39, 147-53 (1960).
11. P. B. Hirsch, *Proc. Roy. Soc.* A226, 143 (1954).
12. H. W. Sternberg, C. L. Delle Donne, R. E. Markby and I. Wender, *Adv. in Chem. Series*, Vol. 55, Coal Science (1966) pp. 516-20.
13. H. W. Sternberg, C. L. Delle Donne, R. E. Markby and I. Wender, *Fuel* (London) 45, 469-82 (1966).
14. H. W. Sternberg, R. E. Markby, C. L. Delle Donne and I. Wender, *Bur. Mines Report* 7017 (1967).
15. H. W. Sternberg, R. E. Markby, I. Wender and D. M. Mohilner, *J. Electrochem. Soc.* 113, 1060-62 (1965).
16. H. W. Sternberg, R. E. Markby, I. Wender and D. M. Mohilner, *J. Amer. Chem. Soc.* 89, 186-7 (1967).

17. H. W. Sternberg, C. L. Delle Donne and I. Wender, *Fuel* (London) 47, 219-22 (1968).
18. W. Rüdorf, *Adv. in Inorg. Chem. and Radiochem.* 1, 223-66 (1959).
19. G. R. Henning, *Prog. Inorg. Chem.* 1, 125-205 (1959).
20. A. R. Ubbelohde and F. A. Lewis, Graphite and Its Crystal Compounds, Oxford, Clarendon Press (1960).
21. Yu. Novikov and M. Vol'pin, *Russ. Chem. Rev. (English trans.)* 40, 733 (1971).
22. L. B. Ebert, *Ph.D. Dissertation*, Stanford University (1975).
23. S. D. James, *J. Electrochem. Soc.* 122, 921-27 (1975).
24. L. B. Ebert, R. A. Huggins and J. I. Brauman, *Bull. Amer. Phys. Soc.* 18 1578 (1973).
25. F. R. Gamble, J. H. Osiecki, M. Cais, R. Pisharody, F. J. DiSalvo and T. H. Geballe, *Science* 174, 493-7 (1971).
26. A. D. Yoffe, *Annual Review of Materials Science* 3, 147-50 (1973).
27. F. R. Gamble and T. H. Geballe, Treatise on Solid State Chemistry, Ed. N. B. Hannay, Vol. 2, Plenum Press (1975).

SYNTHESIS OF LARGE DIAMONDS FOR OPTICAL COMPONENTS

J. J. Gilman

ABSTRACT

A synthetic process currently exists that produces gem (or optical) quality diamonds. However, the cost of its product is excessive because it requires massive equipment which leads to a high ratio of capital investment charges to product throughput value. Superior materials for constructing the high pressure chamber might change this situation significantly by allowing a redesign. Are such materials feasible?

From the theory of the plastic stability of a pressurized spherical hole plus the theory of indentation hardness, it can be deduced that a candidate material must have a hardness number of about 450 kg/mm^2 at 1350° . Very few substances can satisfy this condition. In addition, a candidate material must resist carburization at 1350°C and 60 kbars of pressure; and not be soluble in the transition metal alloy that is used as a catalytic flux. A survey of candidates is reported here.

SYNTHESIS OF LARGE DIAMONDS FOR OPTICAL COMPONENTS

J. J. Gilman

Diamond crystals have highly desirable optical, mechanical, and thermal properties. They transmit light over a wide band of wavelengths, they have the greatest hardness (strength) among known substances, and they have the highest known thermal conductivity at room temperature. A serious deficiency is their high cost.

A synthetic process currently exists that produces diamonds of optical quality (Strong and Wentorf, 1972). It is based on the dissolution of small diamonds in an iron-nickel flux and precipitation of the dissolved carbon at a cooler position in the melt. The required temperature and pressure are about 1400°C and 57 kbar, respectively. Production of a one carat diamond requires about one week of growth time (R. H. Wentorf, 1971). This very low production rate using expensive equipment, skilled labor, and specialized materials leads to manufacturing cost in the 10^2 - 10^3 \$/carat range which is comparable with the cost of natural diamonds and prohibitive for optical or mechanical components.

In order to substantially decrease the manufacturing cost for large diamonds a substantially different process will be necessary. This in turn will require the use of superior

materials for the reaction chamber. Such materials are not easily found and may not even exist. The purpose of this study has been to survey the properties of candidate materials in order to evaluate the feasibility of a new process.

The present process is limited by the marginal strength of cemented-WC under the operating conditions. This makes it necessary to provide massive supporting structures for the carbide dies and leads to relatively rapid die failure. A substantially stronger material would allow larger, lighter, and more durable reaction chambers to be designed and built.

Materials Requirements

A candidate material for the reaction chamber must have at least the following attributes:

- a. high strength at 1300-1400°C
- b. stability against carburization
- c. immiscibility in the flux

The required strength can be estimated from the theory of the plastic stability of a thick spherical shell under internal pressure (Hill, 1950). The pressure that can be withstood is given by:

$$P = \frac{2Y}{3} \left[1 + \ln \left(\frac{2E}{3Y} \right) \right]$$

where Y is the compressive yield stress and E is Young's modulus. For a hard solid, $H = G/6 = 2Y$ (Gilman, 1975). If Poisson's ratio = 1/4, $G = 0.4E$ (or $E/Y = 30$) the equation above yields:

$$P \approx 2.7Y \approx 1.35H$$

Taking the required pressure for diamond growth to be 60 kbar this leads to a required hardness value of:

$$H = 450 \text{ kg/mm}^2$$

and this hardness value refers to a temperature of at least 1350°C.

Candidate Materials

Very few materials can satisfy the strength requirement above and it is probably a minimum requirement. It is unlikely that any polycrystalline material can perform better than the presently used cemented-WC because of grain boundary weakness. Therefore, proposed candidates will have monocrystalline form. The estimated hardnesses of a number of high temperature materials are listed in Table I. A few are not present such as $Y_3Al_5O_12$ and other garnets because data could not be found for them. According to Fitzgerald (1963), B_4C loses very little hardness up to 1400°C, but he used a re-bound method for his measurements so they are not directly comparable with the other values.

Asterisks have been placed after the qualifying candidates in Table I. There are relatively few of them. Although it appears to have adequate hardness, boron must be eliminated because it dissolves in diamond. The best of the remaining candidates appear to be:

Table I. Hot Hardness of Various Refractory Materials

Material	H at 1350°C (kg/mm ²)	References
Boron	1000*	Westbrook (1966)
BeO	60	"
BeAl ₂ O ₄	730*	"
Al ₂ O ₃	350	"
Y ₂ O ₃	150	"
TiO	280	"
TiO ₂	40	"
ZrO ₂	95	"
MgAl ₂ O ₄	660*	"
MgO	40	"
NiO	25	"
MoSi ₂	300	"
ZrSiO ₄	20	"
TiB ₂	450*	"
Fe ₄ Mo ₂ C	700*	"
MoRe (sigma)	620*	"
HfB ₂	700*	Sanders & Probst (1966)
ZrB ₂	310	"
NbB ₂	310	"
LaB ₆	500*	"
HoB ₄	230	"
TiC	4	Hollox (1969)
VC	11	"
VC + 25% TiC	75	"
UC	30	Henry et al. (1970)
W ₂ B ₅	310	Koester & Moak (1967)
ZrB ₂	320	"
HfB ₂	360	"
NbC	200	"
(W _{6.5} Cr _{1.4} Re _{1.4} Ta ₇) ₂ C	450*	"
W ₂ C	300	"
TaC	350	"
HfC	200	"
SiC	1300(?)*	Levitt (1970)
WC	450*	Westbrook & Stover (1967)
ZrC	90	"
B ₄ C	~4000	Fitzgerald (1963)

- a. SiC
- b. BeAl_2O_4 (chrysoberyl)
- c. $\text{Fe}_4\text{Mo}_2\text{C}$
- d. HfB_2
- e. B_4C

Of these, only BeAl_2O_4 has been made in the form of large crystals. It should be stable against reduction by carbon since both MgO and Al_2O_3 are quite stable according to the charts of Richardson and Jeffes (1948); data for BeO are not available.

The solubilities of candidates a, c, d and e in nickel or nickel-iron fluxes might disqualify them, but the author is not aware of data relevant to this.

Hot hardness data for SiC were not found in the literature so the value given in Table I was estimated from tensile and bend test data. Even if SiC had adequate properties it is so difficult to make in the form of large monocrystals that it would not be chosen unless no other material could be found.

In view of the good hot hardness of chrysoberyl, other complex oxides such as yttrium aluminum garnet might be candidate materials, but data are not available for them.

Conclusions

The data that have been collected indicate that some materials (other than those in current use) have the intrinsic strength that is required in constructing equipment for diamond

synthesis. However, a development and evaluation program will be necessary as a first step toward their use in an improved process. They must be produced in monocrystalline form and their compatibility with carbon and transition metal fluxes must be evaluated.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

REFERENCES

1. L. M. Fitzgerald, "The Hardness at High Temperatures of Some Refractory Carbides and Borides," *J. Less-Common Met.* 5, 356 (1963).
2. J. L. Henry, R. Blickensderfer, D. Paulson, "Hot Hardness, Thermal Diffusivity, and Electrical Resistivity of Uranium Oxycarbides," *J. Amer. Cer. Soc.* 53, 335 (1970).
3. R. Hill, The Mathematical Theory of Plasticity, Oxford Press, 97ff (1950).
- G. Hollox, Mechanical and Thermal Properties of Ceramics, U. S. Nat. Bur. Stand., Spec. Publ. #303, p. 201, Ed. by J. B. Wachtman (1969).
4. J. J. Gilman, "Relationship Between Impact Yield Stress and Indentation Hardness," *J. Appl. Phys.* 46, 1435 (1975).
5. A. P. Levitt (Editor), Whisker Technology, p. 157, Wiley-Interscience, New York (1970).
6. F. D. Richardson and J. H. E. Jeffes, "The Thermodynamics of Substances of Interest in Iron and Steel Making from 0°C to 2400°C," *J. Iron and Steel Inst.* 160, 261 (1948).
7. W. A. Sanders and H. B. Probst, "Hardness of Five Borides at 1625°C," *J. Amer. Cer. Soc.* 49, 231 (1966).
8. H. M. Strong and R. H. Wentorf, "The Growth of Large Diamond Crystals," *Die Naturwissenschaften* 59, 1 (1972).
9. R. H. Wentorf, "Some Studies of Diamond Growth Rates," *J. Phys. Chem.* 75, 1833 (1971).
10. J. H. Westbrook, "The Temperature Dependence of Hardness of Some Common Oxides," p. 47 in Proprietes des Materiaux a Haute Temperature, Publ. by Centre National de la Recherche Scientifique, Paris (1966).
11. J. H. Westbrook and E. R. Stover, High Temperature Materials and Technology, Ed. by Campbell and Sherwood, p. 312, J. Wiley, New York (1967).

ON THE DESIGN OF ALLOYS FOR HIGH STRFNGTH
AND FRACTURE TOUGHNESS

M. Cohen

ABSTRACT

It is postulated that fracture toughness increases with the energy that is absorbed in the plastic zone at the tip of a moving crack, and that this absorbed energy increases with the strain-hardening rate ($d\sigma/d\epsilon$). When the yield strength is raised, as in the martensitic strengthening of steel, the energy absorbed by the plastic zone is decreased because the critical size of the plastic zone and the strain hardening within it are reduced. However, by starting with metastable medium-carbon austenitic steels (intially warm worked to provide high yield strength), the base composition can be chosen to favor a strain-induced martensitic transformation in the plastic zone at a crack tip; and inasmuch as the martensite thus generated is significantly stronger than the parent austenite, the effective strain-hardening rate is very high. TRIP steels make use of this phenomenon, and the observed combinations of strength and toughness are probably the best that have been attained in any material to date ($\text{Y.S.} \approx 250 \text{ ksi}$ and $K_{Ic} \approx 100 \text{ ksi}\sqrt{\text{inch}}$).

Unfortunately, the strain-induced nucleation of martensite is usually quite temperature-sensitive, and adiabatic heating due to high strain rates raises the temperature

sufficiently to render inoperative this special mode of strain hardening. We then inquire into the possibility of avoiding the latter undesirable feature by alloy design.

Recent work has shown (G. B. Olson and Morris Cohen, "Kinetics of Strain-Induced Martensitic Nucleation," *Met. Trans.* 6A (April 1975) 791-795) that, based on a shear-band intersection mechanism, the volume fraction of martensite (f^M) resulting from strain-induced nucleation at such intersection sites can be expressed as:

$$f^M = 1 - \exp\{-\beta[1 - \exp(-\alpha\epsilon)]^n\}$$

where ϵ is the plastic strain, n is a fixed exponent ≈ 2 for randomly oriented shear bands, α is a physical parameter depending on the stacking-fault energy ($\Delta G^{fcc \rightarrow hcp}$) and measures the propensity to form shear bands during plastic deformation, and β is a physical parameter depending on the driving force for the martensitic transformation ($\Delta G^{fcc \rightarrow bcc}$). Since both ΔG 's are known as a function of composition and temperature, it becomes possible to choose compositions which will not only keep f^M vs. ϵ in the right regime, but will also minimize the temperature-dependence of this functional relationship. The indications are that the replacement of nickel by manganese in the steels under consideration will meet the necessary conditions.

ON THE DESIGN OF ALLOYS FOR HIGH STRENGTH
AND FRACTURE TOUGHNESS

M. Cohen

I. Introduction

The energy absorbed in the plastic zone at the tip of a moving crack increases with the plastic zone size and, in turn, with the rate of strain hardening. It usually happens, however, that the strengthening of alloys by various methods results in a reduced rate of strain hardening at the higher stress levels which would take advantage of the enhanced strength. Thus, K_{IC}/Y ratio tends to decrease with increasing Y , where K_{IC} is the plane-strain fracture toughness and Y is the 0.2% offset yield strength. This inverse trend between K_{IC} and Y places severe limits on the engineering usefulness of high-strength steels, and reduces their tolerance for flaws.

Some typical data for high-strength steels are given in Table 1. It will be noted that there is a general decrease in the plastic-zone size with increasing strength level; this is shown most clearly in the maraging-steel data where an appreciable range of strength levels has been tested.

TRIP steels, on the other hand, undergo a strain-induced transformation in the plastic zone at a crack tip, and hence are capable of an especially high rate of strain harden-

TABLE 1. Some Strength/Toughness Data for High-Strength Steels
(Longitudinal Properties)

<u>Steel</u>	<u>Y (ksi)</u>	<u>K_{Ic} (ksi$\sqrt{\text{in}}$)</u>	<u>K_{Ic}/Y</u>	<u>Plastic-Zone Size (in)*</u>	<u>Ref.</u>
9Ni-4Co	180	145	0.81	0.10	1
	184	97	0.53	0.04	1
	184	66	0.36	0.02	1
	184	135	0.73	0.09	1
	184	113	0.61	0.06	1
	Avg. 183	111	0.61	0.06	
18 Ni Maraging	225	104	0.46	0.03	1
	226	102	0.45	0.03	1
	233	94	0.40	0.03	1
	234	69	0.29	0.01	1
	242	85	0.35	0.02	2
	259	68	0.26	0.01	2
	285	52	0.18	0.005	2

$$* \text{Plastic-zone size} \approx \frac{1}{2\pi} \left(\frac{K_{Ic}}{Y} \right)^2$$

ing. These steels are alloyed to be fully austenitic (fcc) at room temperature, and can be strengthened to levels of $Y \approx 250$ ksi by warm rolling to approximately 75% reduction in thickness at about 450°C . Effectively, then, this austenite is in its strain-hardened condition, and its rate of strain hardening in a tensile test at room temperature should be of the order of $d\sigma/d\varepsilon = 200$ ksi. But actually, the observed rate of strain hardening is several times that expected, the anomaly being due to the strain-induced transformation of the austenite to martensite (bcc) in the plastic zone. Consequently, we may have a case here in which higher levels of K_{Ic}/Y can be attained because the strain hardening that occurs at the crack tip is not simply a continuation of the prior strain-hardened state of the parent phase, but results from a new process that comes into play when the fracture toughness is being tested.

Fortunately, strain-induced martensitic transformations have already been studied in considerable detail, quite separately from the fracture problem.

II. Strain-Induced Martensitic Nucleation

If the stacking-fault energy of an austenitic steel is sufficiently low, faulting occurs during plastic deformation to generate shear bands which can be modeled as hcp bands in the fcc matrix. It is found^{3,4} that the intersection of such bands provides favorable sites for the nucleation of austenite to martensite (bcc), and so the kinetics of the deformation-

induced transformation can be related to the thermodynamic driving forces to produce hcp bands ($\Delta G^{fcc \rightarrow hcp}$) and then to nucleate martensite ($\Delta G^{fcc \rightarrow bcc}$) at the intersections. With such considerations, the volume fraction of martensite (f^M) as a function of plastic strain (ϵ) can be expressed as:⁵

$$f^M = 1 - \exp\{-\beta[1 - \exp(-\alpha\epsilon)]^n\} \quad (1)$$

where n is a fixed exponent ≈ 2 for randomly oriented shear bands; α is a strain-independent constant representing the rate of shear-band formation with respect to strain at low strains ($df^{sb}/d\epsilon$); and β is proportional to the probability that a shear-band intersection will cause a nucleation event. Both α and β have physical meaning, with their temperature dependences being controlled respectively by the chemical free-energy differences $\Delta G^{fcc \rightarrow hcp}$ and $\Delta G^{fcc \rightarrow bcc}$, which are known as functions of temperature and composition.

Figure 1 indicates how well Eq. (1) fits the experimental data for 304 austenitic stainless steel.⁶ The steel composition has to be carefully adjusted so that martensite will not be present before the plastic straining and yet will form in significant amounts during the straining. This balance is controlled by the α and β parameters. However, for the usual TRIP steels, α and β are temperature-sensitive near room temperature, and a rise of 50°C (caused by adiabatic heating when tested at high strain rates) will inhibit the

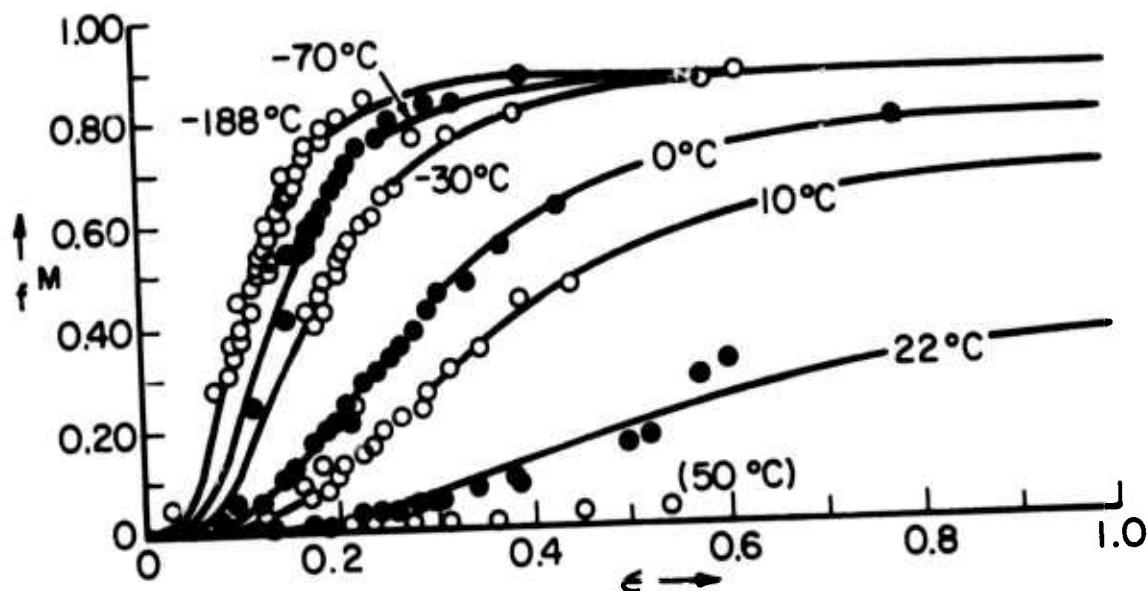


Figure 1. Comparison of calculated transformation curves with data of Angel⁶ for a 304 stainless steel. Experimental data are indicated by points. The solid curves represent the best fit of the derived expression for each temperature, consistent with the assumed temperature dependence of the α and β parameters. (Taken from Ref. 5.)

deformation-induced martensitic transformation, thereby eliminating the extra strain hardening being sought.

From a knowledge of the way that $\Delta G^{fcc \rightarrow hcp}$ and $\Delta G^{fcc \rightarrow bcc}$ change with temperature, it appears that the temperature-dependence of Eq. (1) near room temperature can be minimized by replacing nickel by manganese in the alloy.

III. Strength and Fracture Toughness of TRIP Steels

Relatively little information is available thus far on the K_{IC} values of TRIP steels. Although the nominal composition of this alloy is 9Cr, 8Ni, 4Mo, 2Mn, 2Si, 0.3C, the best data for present purposes happen to be for a TRIP steel containing 0.6% C,⁷ which is probably much too high for the best combination of strength and fracture toughness. Figure 2 shows how K_{IC} varies with test temperature at a strength level of $\gamma = 250$ ksi. The striking change in behavior just above room temperature denotes a transition from the lower temperature regime where the strain-induced transformation is operative, to the higher temperature regime where the transformation does not occur. The loss in K_{IC} due to the absence of the transformation appears to be about 58 $\text{ksi}^{\sqrt{\text{inch}}}$ (extrapolated back to room temperature); i.e., without the transformation, the expected K_{IC} at room temperature would be only 37 $\text{ksi}^{\sqrt{\text{inch}}}$. Roughly speaking, then, the phenomenon under discussion seems to account for more than a doubling of the fracture toughness of the parent phase.

The observed K_{IC} average of 95 ksi $\sqrt{\text{inch}}$ at room temperature in Fig. 2 leads to a K_{IC}/Y ratio of 0.38 and a plastic zone size of 0.02 in., all of which compares favorably with the maraging steel values in Table 1, interpolated to $Y = 250$ ksi. Moreover, by lowering the carbon content to 0.3% in the TRIP steel (instead of the 0.6% C for Fig. 2), considerable improvement in the K_{IC} level may be anticipated.

Figure 2 also illustrates the marked temperature dependence of K_{IC} in this material. With the previously suggested compositional changes to minimize the temperature sensitivity of the strain-induced transformation, this type of high-strength steel might also exhibit attractive fracture toughness even under conditions of dynamic loading.

Acknowledgement

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAH 15-71-C-0253 with The University of Michigan.

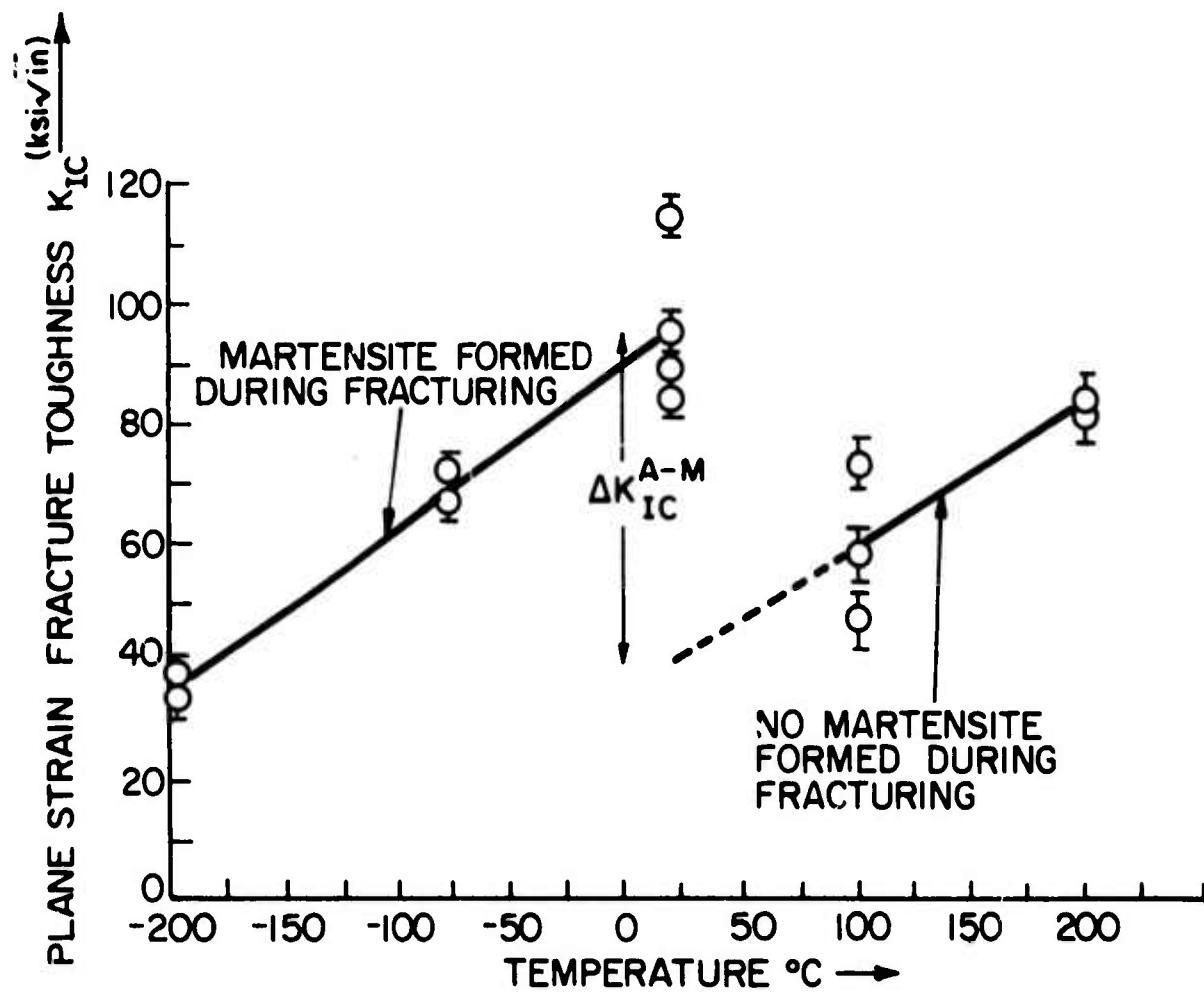


Figure 2. Toughness vs. temperature for alloys deformed 75 pct at 460°C . TRIP steel with 0.6% carbon. (Taken from Ref. 7.)

References

1. Report No. AFML-TR-66-340 (Contract AF33-(657)-11277). Systems Engineering Group, Air Force Systems Command, Wright-Patterson Air Force Base, Dayton, Ohio. (1966)
2. W. F. Brown and J. E. Srawley, ASTM, STP 410 (1966) 1.
3. J. A. Venables, Phil. Mag. 7 (1964) 35.
4. G. B. Olson and M. Cohen, J. Less-Common Metals 28 (1972) 107.
5. G. B. Olson and M. Cohen, Met. Trans. 6A (Apr. 1975) 791.
6. T. Angel, J. Iron and Steel Inst. 177 (1954) 165.
7. S. D. Antolovich and B. Singh, Met. Trans. 3 (1971) 2135.

CONCENTRATION FLUCTUATIONS
IN ADSORBED LAYERS

Brian Bell*, R. Gomer and H. Reiss

ABSTRACT

The temperature and coverage dependence of the mean square concentration fluctuations in a small open domain of an adsorbed layer is discussed for various situations. It is shown that fluctuations decrease with increasing temperature and reach a limiting value when attractive interactions predominate, if a single non-ideal two-dimensional phase exists. Deviations from ideal gas behavior are strongest at half coverage. At very low coverage (low particle concentration) and very high coverage (low hole concentration) ideal behavior is approached.

If the layer consists of a two phase system, for instance a two-dimensional liquid or solid in equilibrium with a two-dimensional gas, fluctuations far below the critical temperature are dominated by fluctuations in the partition between phases. As the critical temperature is approached fluctuations first decrease because the mean concentrations in the two phases approach each other, and then increase very sharply near T_c . Detailed calculations for the single phase situation are given for several approximations: a dilute gas;

*The James Franck Institute, The University of Chicago.

a mean field approximation; a lattice gas in both the Bragg-Williams and the Bethe-Peierls-Weiss approximations. The latter which takes some account of correlations between adsorbate particle positions seems to explain reasonably the presently available experimental observations on chemisorbed layers.

CONCENTRATION FLUCTUATIONS IN ADSORBED LAYERS

Brian Bell, R. Gomer and H. Reiss

In a previous publication¹ it was shown that the time correlation function of current fluctuations in field emission can be utilized for the determination of surface diffusion coefficients of adsorbates on single planes of field emitters. Explicit expressions for the correlation function were worked out with the assumption of no interactions between the adsorbate particles. The current correlation function for $t = 0$, is the mean square fluctuation and is proportional to the absolute mean square adsorbate concentration fluctuation, $\langle \Delta n \rangle^2$

$$\langle \Delta n \rangle^2 \equiv \langle n^2 \rangle - \langle n \rangle^2 \quad (1)$$

For noninteracting adsorbate particles

$$\langle \Delta n \rangle^2 = \langle n \rangle \quad (2)$$

where $\langle n \rangle \equiv \bar{n}$ is the mean number of adsorbate particles in the region of observation.

Experiments by J. Chen and R. Gomer² indicate that strong deviations from such ideal behavior can occur. For O or CO on the (110) plane of W, for instance, $\langle \Delta n \rangle^2$ first increases with T and then levels off (Fig. 1). The present paper explores various possible deviations of $\langle \Delta n \rangle^2$ from ideal

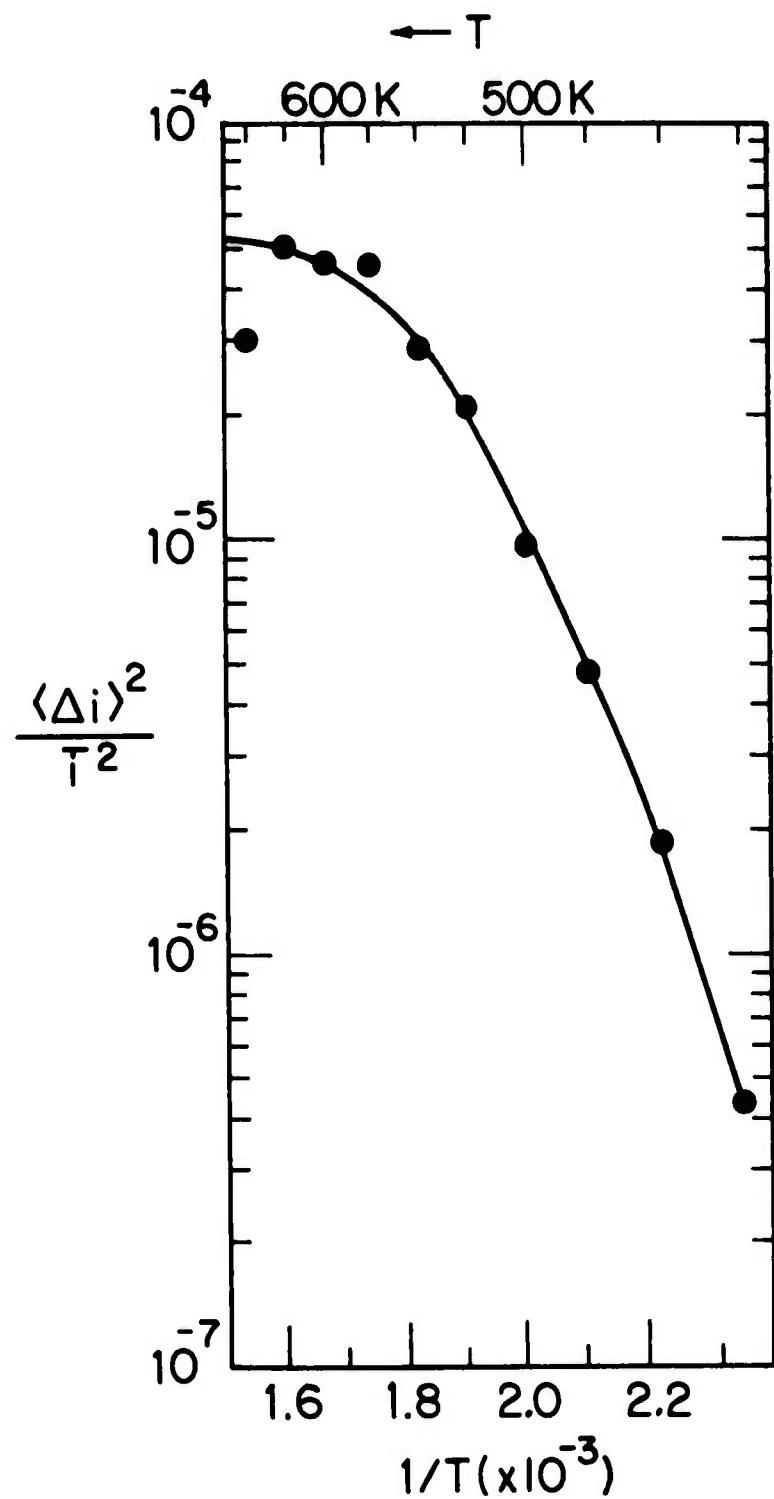


Figure 1. Relative mean square current fluctuation in field emission from an oxygen covered (110) plane of a tungsten field emitter vs. $1/T$. (Ref. 2.) The radius of the probed region is roughly 50\AA ; coverage $\theta \sim 0.4$. The mean square current fluctuation is proportional to the absolute concentration fluctuation.

gas behavior that might be encountered in chemisorption or physisorption.

It can be shown from general considerations³ that

$$\langle \Delta n \rangle^2 / \bar{n} = (\bar{n}/A) K kT \quad (3)$$

where A is the area of the probed region and K the two-dimensional compressibility

$$K = -(1/A) \partial A / \partial p_s \quad (4)$$

with p_s the two-dimensional pressure. By analogy to p-V diagrams for 3-dimensional phases K decreases with increasing T if there are short range repulsive and long range attractive interactions. Thus an increase of K with T implies the predominance of repulsive interactions at all separations. This could come about if dipole-dipole repulsions predominate.

I. Single Phase of an Imperfect Gas

We shall now consider the adsorbate as an imperfect two-dimensional gas in various approximations. The first to be presented, the dilute gas and mean field approximations, ignore the structure of the substrate completely, and are therefore most appropriate to situations where binding is relatively non-specific, for instance physisorption. Next, lattice gas approximations, which emphasize the existence of discrete binding sites will be discussed. These probably apply to most cases of chemisorption.

Dilute Gas

An equation of state can be obtained by standard methods⁴ and then leads to

$$\langle \Delta n \rangle^2 / \bar{n} = (1 - \theta \cdot I)^{-1} \quad (5)$$

where $\theta = \bar{n}/A$ and

$$I = 2\pi \int_0^{\infty} (e^{-u\beta} - 1) r dr \quad (6)$$

Here $\beta = 1/kT$, and $u(r)$ is the pair potential, r being the separation between adsorbate particles. If $u(r)$ is taken to be

$$u(x) = \infty, \quad x < 1. \quad (7)$$

$$u(x) = u_0 x^{-s}, \quad x \geq 1 \quad (8)$$

with $x = r/r_0$, r_0 being the distance of closest approach, one obtains

$$\langle \Delta n \rangle^2 / \bar{n} = [1 + \theta \pi r_0^2 (1 - 2 \int_1^{\infty} (e^{-u_0 \beta x^{-s}} - 1) x dx)]^{-1} \quad (9)$$

If $u_0 < 0$, corresponding to long range attractive forces, the integral in Eq. (9) is positive and approaches 0 as $T \rightarrow \infty$. Thus $\langle \Delta n \rangle^2$ decreases at T increases. On the other hand if $u_0 > 0$, i.e., if the potential is everywhere repulsive, the integral will be negative, approaching 0 as T increases, and thus $\langle \Delta n \rangle^2$ will increase and then level off as T increases. Theoretical considerations⁵ indicate that a part of $u(r)$ is in

fact oscillatory for adsorbed particles because of substrate-mediated interactions. Nevertheless the above result is not altered: the repulsive part of the potential must dominate, even at long range, to account for an increase in $\langle \Delta n \rangle^2$ with increasing T. Table I lists numerically computed values of

$$I' = \int_1^{\infty} (e^{-(u_0/kT)x^{-3}} - 1) x dx \quad (10)$$

for various positive values of u_0/kT , corresponding to repulsive interaction. The exponent $s = 3$ corresponds to dipole-dipole repulsion. Fig. 2 shows a plot of $\langle \Delta n \rangle^2/\bar{n}$ vs. u_0/kT for $\theta\pi r_0^2 = 1$, i.e., high adsorbate coverage, for which Eq. (9) is not expected to be quantitatively applicable, having been derived specifically for the dilute gas case.

Mean Field

It is also possible to treat the case of high adsorbate concentration in the mean field approximation. We assume that the effective potential energy U felt by an adsorbate particle because of the presence of all other adsorbate is given by

$$U = 2\pi q\alpha P n / A = \alpha q \Delta \phi \quad (11)$$

where q is the adsorbate charge, P the dipole moment per adsorbate particle and α a coverage dependent factor⁶ ≤ 1 given by

$$\alpha = V(d_0) / V(\infty) = V(d_0) / \Delta \phi \quad (12)$$

where V is the dipole layer potential, d_0 the adsorbate-surface

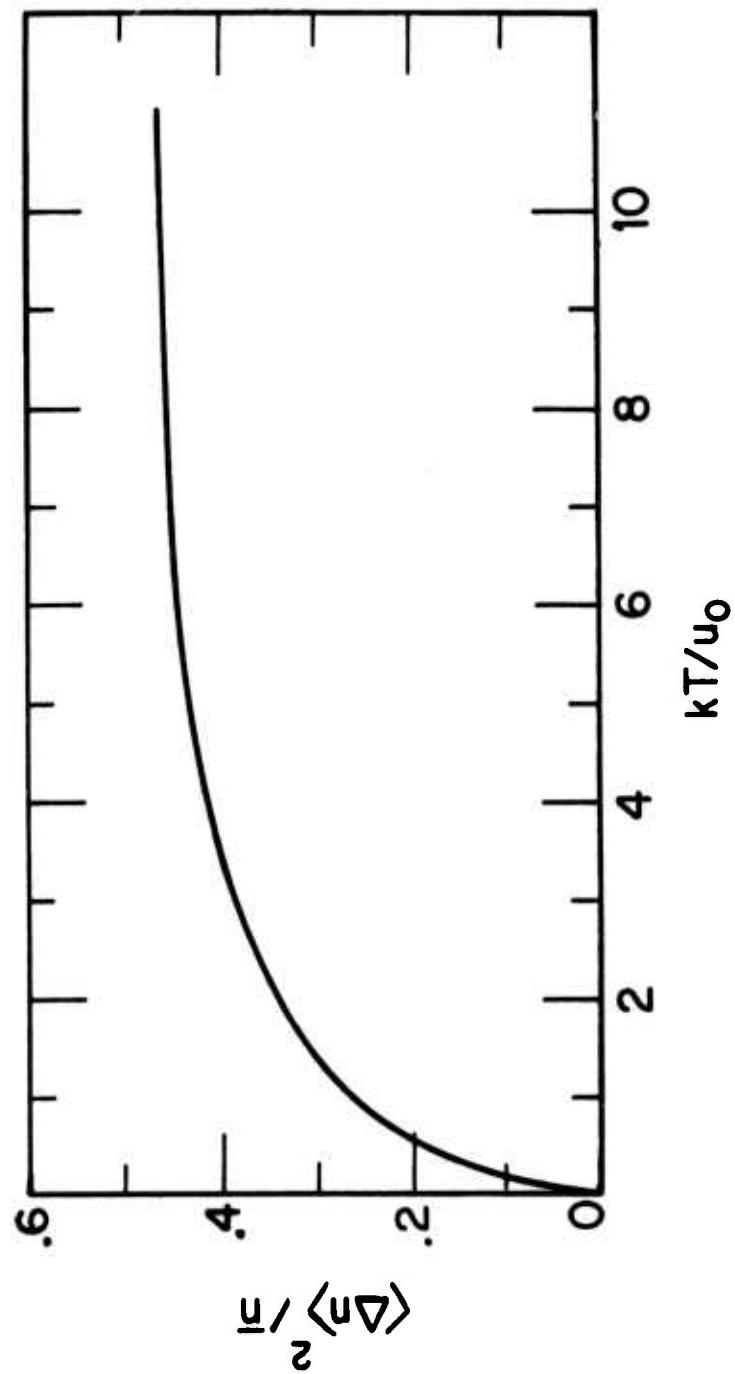


Figure 2. $\langle \ln n^2/n \rangle$ vs. kT/u_0 for the dilute gas approximation, Eq. (9)
 $\theta \pi r_0^2 = 1$.

separation and $\Delta\phi$ the work function change caused by the ad-layer.

The partition function is given approximately⁴ by

$$\ln Z = \bar{n} \ln(A - A_O) - \bar{n}U/kT \quad (13)$$

where $A_O = (1/2)\bar{n}\pi r_O^2$ is the excluded area, so that the two-dimensional pressure p_s becomes

$$p_s = kT(\partial \ln Z / \partial A) = \bar{n}kT/(A - A_O) + \bar{n}^2 \alpha q 2\pi P/A^2 \quad (14)$$

and K is given by

$$K = \frac{1}{\frac{A_O kT}{(A - A_O)^2} + \frac{4\bar{n}^2 \pi P \alpha q}{A^2}}$$

The

$$\begin{aligned} \langle \Delta n \rangle^2 / \bar{n} &= \frac{1}{\left(\frac{1}{1 - \frac{1}{2} \theta \pi r_O^2} \right)^2 + 2(\theta/kT) 2\pi P \alpha q} \\ &= \frac{(1 - \theta \pi r_O^2/2)^2}{1 + (1 - \theta \pi r_O^2/2)^2 (\theta/kT) 4\pi P \alpha q} \end{aligned} \quad (16)$$

Eq. (16) shows that in the mean field approximation $1/\langle \Delta n \rangle^2$ varies inversely with T and linearly with coverage for a gas of point particles. The finite size of the adsorbate tends to decrease fluctuations as θ increases. This corresponds to the fact that one may consider fluctuations to involve particles at low θ and holes at high θ . Although the

diffusion coefficient of holes may differ from that of particles, a dilute "hole gas" should again be nearly perfect, just like a dilute particle gas. Eq. (16) contains this feature qualitatively. If we define maximum coverage $\theta_m = 2/\pi r_o^2$ the number of holes is given by

$$n_h = n_m - n = n(n_m/n - 1) = n(\theta_m/\theta - 1) \quad (17)$$

where n_m is the maximum number of adsorbate particles. Consequently,

$$\langle \Delta n \rangle^2 \rightarrow n(\theta/\theta_m)^2 (\theta_m/\theta - 1)^2 = n_h (\theta/\theta_m)^2 (1 - \theta/\theta_m) \quad (18)$$

as $\theta \rightarrow \theta_m$. It is not entirely obvious why this expression contains the extra factor $(\theta/\theta_m)^2 (1 - \theta/\theta_m)$. A plot of $\langle \Delta n \rangle^2 / \bar{n}$ vs. T based on Eq. (16) is shown in Fig. 3.

Lattice Gas

Possibly a somewhat better approximation, at least for chemisorption, may be provided by the lattice gas model. Even in its simplest form this gives the expected linear behavior with concentration both for particles and holes. We start by assuming that the Helmholtz free energy, F , of a lattice gas is given by

$$F = \frac{z\epsilon}{2} nx + kT[n \ln x + m(1-x) \ln(1-x)] \quad (19)$$

Here ϵ is the nearest neighbor interaction energy (all others being neglected), z is the number of nearest neighbors of a site; m is the number of sites, n the number of particles dis-

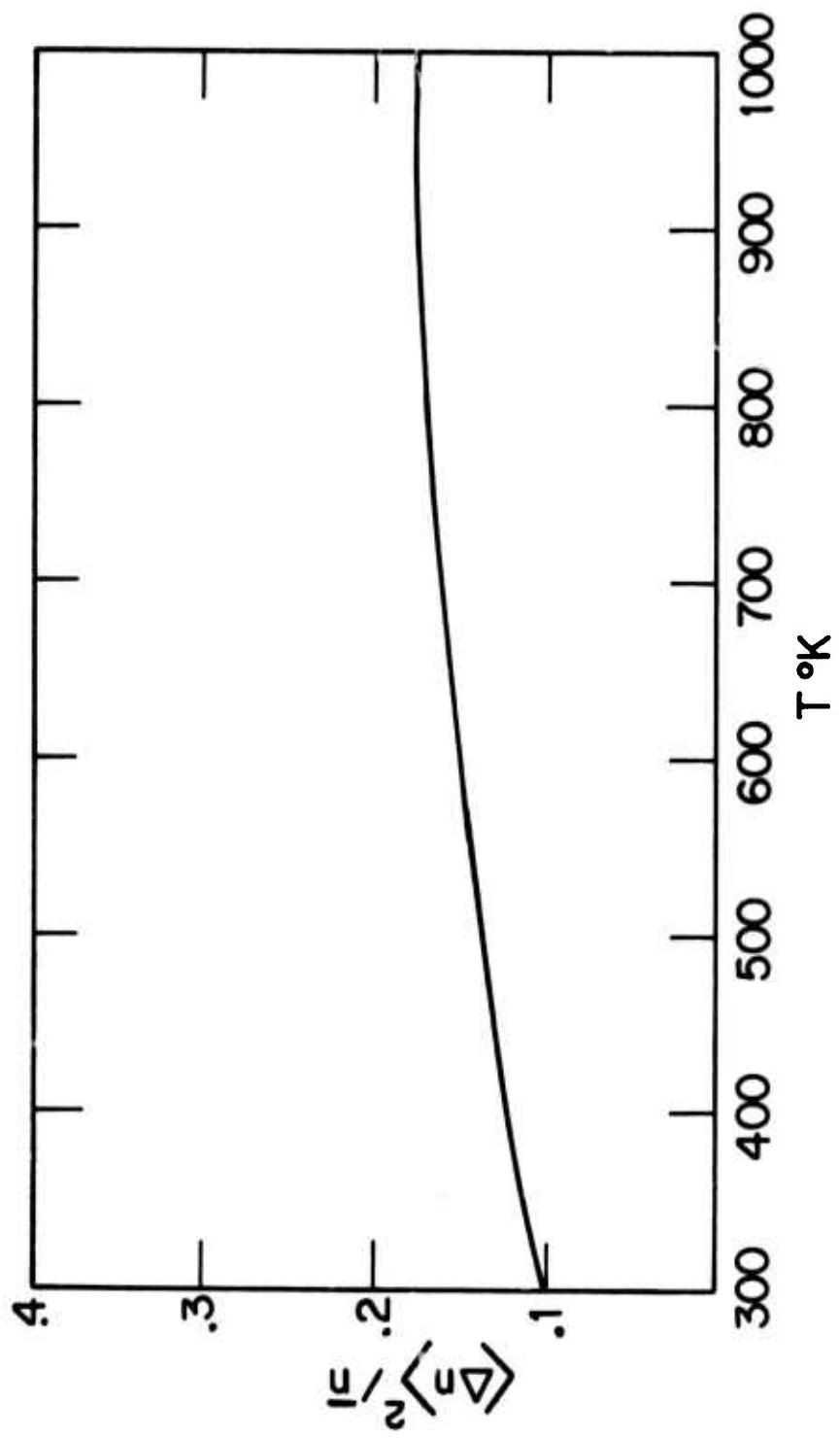


Figure 3. $\langle \Delta n \rangle^2 / \bar{n}$ vs. T for the mean field approximation, Eq. (16). For specificity the following values have been assumed: $\Delta\phi = 1$ eV; $q = 0.1$ e $^-$; $\alpha = 0.7$; $\theta\pi r_0 = 1$.

tributed on the m sites and $x = n/m$ the fraction of filled sites. In this approximation we ignore the fact that the existence of an interaction leads, at least for $\epsilon \gg kT$, to a non-random distribution of nearest neighbor occupancy. Let the area per site be a so that $A = ma$. We then obtain

$$\begin{aligned} p_s &= \frac{-1}{a} (\partial F / \partial m)_n \\ &= (-kT/a) [\ln(1-x) - \eta x^2] \end{aligned} \quad (20)$$

where $\eta = z\epsilon/2kT$, and straightforwardly

$$\langle \Delta n \rangle^2 / \bar{n} = \frac{1-x}{1+2\eta x(1-x)} \quad (21)$$

Eq. (21) has the same T-dependence as Eq. (16), but shows a more plausible coverage dependence: At very small x $\langle \Delta n \rangle^2 = n(1-x) \approx n$ while at very high x it assumes the form $\langle \Delta n \rangle^2 = n_h$. This behavior is insured by the factor $x(1-x)$ in the denominator which approaches 0 as $x \rightarrow 0$ and as $x \rightarrow 1$. We also see that attractive interactions ($\eta < 0$) lead to a decrease in fluctuations with increasing T while repulsive interactions ($\eta > 0$) lead to increasing values which approach $(1-x)$ asymptotically.

For the experimental cases studied so far the fluctuations increase far more rapidly with T than is predicted by any of the models treated so far. It is possible that this results from neglect of correlation, i.e., in the lattice model from the assumption of completely random nearest neighbor

occupancy. It is of course extremely difficult to treat this problem correctly, even in the lattice gas model, but an approximate treatment based on the Bethe-Peierls-Weiss model⁷ is possible. This model considers in detail a pair of adjacent sites, and treats the environment of this pair in the mean field approximation. If again we consider only nearest neighbor interactions the canonical partition function is

$$Z = \sum_y g(n, y) e^{-y\epsilon/kT} \quad (22)$$

where $g(n, y)$ is the number of ways of putting n atoms on m sites, such that there are y pairs of adjacent atoms. The equilibrium number of pairs is \bar{y} and can be found from A in the usual way. We now define an effective number of pairs, y_o , such that

$$Z = e^{-y_o\epsilon/kT} \frac{m!}{n!(m-n)!} \quad (23)$$

The last factor is just the number of ways of putting n atoms randomly on m sites, and the effects of interactions are thus contained in y_o . It can be shown⁷ that

$$\bar{y} = \frac{\partial}{\partial \beta} (\beta y_o) \quad (24)$$

where $\beta = 1/kT$, and further that

$$2\bar{y}(zm-2zn+2\bar{y}) = (an-2\bar{y})^2 e^{-\epsilon/kT} \quad (25)$$

Eq. (25) is quadratic in \bar{y} and has the solution

$$\bar{y} = \frac{zn}{2} \left(1 - \frac{2(1-x)}{1+\alpha} \right) \quad (26)$$

where

$$\alpha = \{1-4x(1-x)(1-e^{-\epsilon/kT})\}^{\frac{1}{2}} \quad (27)$$

Thus we have an expression for \bar{y} , the mean number of pairs, and, in Eq. (24) a relation between \bar{y} and y_o , the effective number of pairs. We can now use Eq. (23) for the canonical partition function, expressed in terms of y_o , to calculate the fluctuations. Proceeding as before, that is finding the compressibility from Z , we obtain

$$\langle \Delta n \rangle^2 = x^2 kT \left[\frac{\partial^2 z}{\partial m^2} \right] \quad (28)$$

which yields, using Eq. (23)

$$\frac{\langle \Delta n \rangle^2}{\bar{n}} = \frac{1-x}{1 + \left(\frac{m}{x} \right) \left(\frac{\partial^2 y_o}{\partial m^2} \right) (1-x)n} \quad (29)$$

We shall evaluate Eq. (29) for three cases: low coverage, $x=.5$, and high coverage.

1. Low Coverage

For $n \ll m$, $\bar{y} \ll n$, Eq. (25) reduces to

$$\bar{y} = \frac{zny}{2} e^{-\epsilon/kT} \quad (30)$$

From Eq. (24) we obtain

$$\beta y_o = \frac{znx}{2} \int_0^{\beta} e^{-\beta' \epsilon} d\beta' \quad (31)$$

or

$$y_O = \frac{znx}{2\beta\varepsilon} (1-e^{-\varepsilon/kT}) \quad (32)$$

Using the result (32) in Eq. (29) finally yields

$$\frac{\langle \Delta n \rangle^2}{\bar{n}} = \frac{1-x}{1+zx(1-x)(1-e^{-\varepsilon/kT})} \quad (33)$$

Eq. (33) is reminiscent of Eq. (21) for the simple lattice gas model and reduces to it for $\varepsilon \ll kT$. A plot of $\langle \Delta n \rangle^2 / \bar{n}$ vs. ε/kT based on Eq. (33) is shown in Fig. 4.

2. $\bar{x} = 0.5$

The general intermediate coverage case is rather messy, but simplifications occur for half coverage, $x = 0.5$. For this case

$$\alpha = e^{-\varepsilon/2kT} \quad (34)$$

$$\frac{\partial \alpha}{\partial m} = 0 \quad (35)$$

and

$$\frac{\partial^2 \alpha}{\partial m^2} = 4(1-e^{-\varepsilon/kT})e^{+\varepsilon/2kT} \left(\frac{\partial \bar{y}}{\partial m} \right)^2 \quad (36)$$

where α is defined by Eq. (27). Differentiating Eq. (26) twice with respect to m yields

$$\frac{\partial^2 \bar{y}}{\partial m^2} = \frac{zx}{m} \left\{ \frac{e^{\varepsilon/2kT}}{1+e^{\varepsilon/2kT}} + \frac{e^{\varepsilon/2kT}(e^{\varepsilon/kT}-1)}{2(1+e^{\varepsilon/2kT})^2} \right\} \quad (37)$$

We can now find $\partial^2 y_O / \partial m^2$ from

$$\frac{\partial^2 y_O}{\partial m^2} = \frac{1}{\beta} \int_0^\beta \frac{\partial^2 \bar{y}}{\partial m^2} d\beta \quad (38)$$

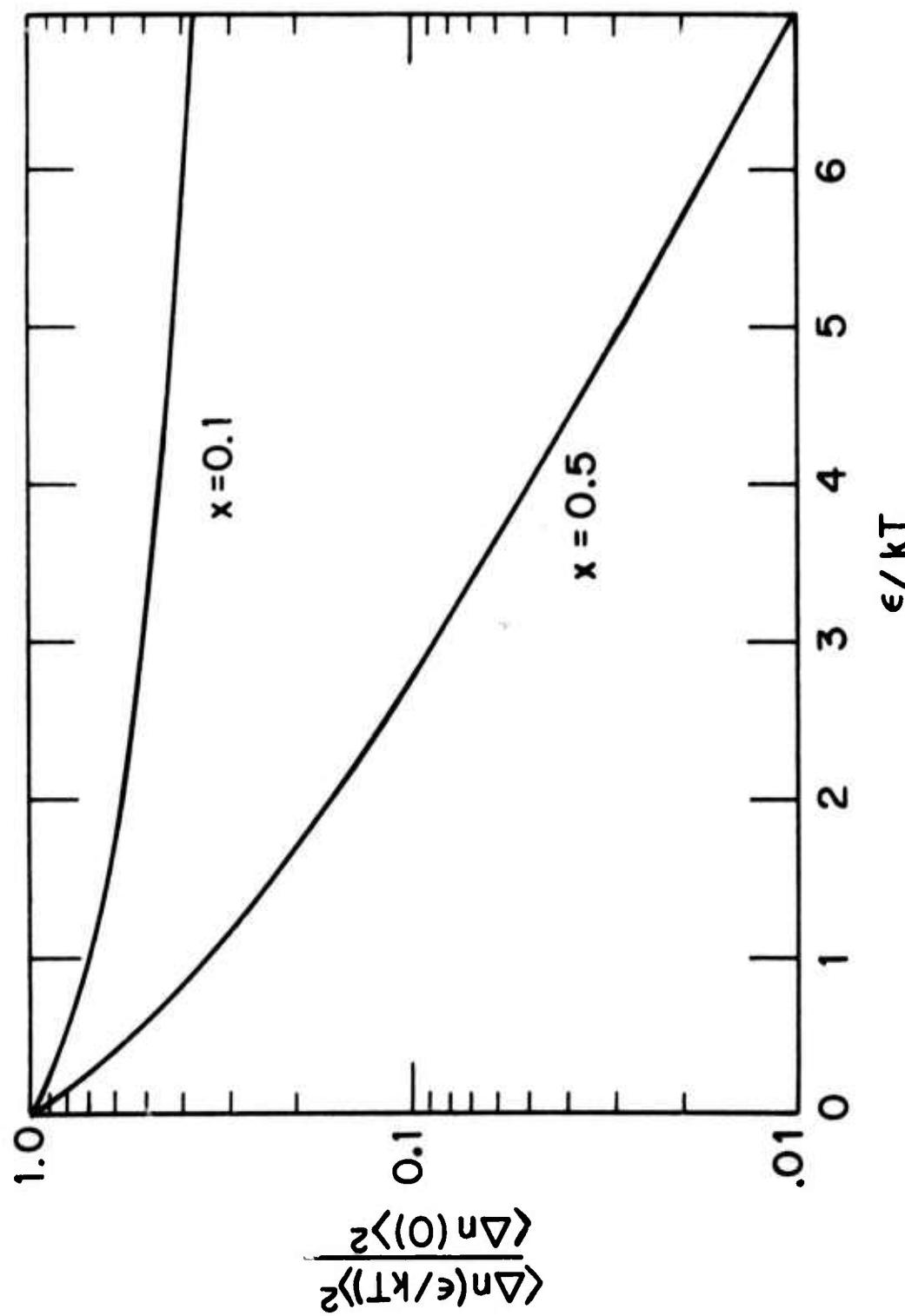


Figure 4. $\langle \Delta n \rangle^2$ as a function of ϵ/kT , divided by the infinite temperature value, $\langle \Delta n(0) \rangle^2$ vs. ϵ/kT , plotted according to Eqs. (33) and (40) for $x = 0.1$ and 0.5 , respectively; $z = 6$ in both cases.

using Eq. (37) for $\partial^2 \bar{y} / \partial m^2$. The result, after some straightforward manipulation, is

$$\frac{\partial^2 \bar{y}_0}{\partial m^2} = \frac{kT}{\varepsilon} \frac{zx}{m} (1 - e^{-\varepsilon/2kT}) e^{\varepsilon/2kT} \quad (39)$$

which leads to

$$\frac{\langle \Delta n \rangle^2}{\bar{n}} = \frac{1-x}{1+z(1-x)e^{\varepsilon/2kT}(1-e^{-\varepsilon/2kT})} \quad (40)$$

where $x = 0.5$, for this special case. Eq. (40) shows the strongest variation of $\langle \Delta n \rangle^2$ with temperature of any of the cases considered so far because of the exponential $e^{\varepsilon/2kT}$ in the denominator. It also reduces to Eq. (21) for $\varepsilon \ll kT$. It is interesting that an intermediate limiting form of Eq. (40) is

$$\frac{\langle \Delta n \rangle^2}{\bar{n}} = \frac{1-x}{1+z(1-x)(\varepsilon/2kT)e^{\varepsilon/2kT}} \quad (41)$$

This is the form one would obtain from the simple lattice gas model by arbitrarily multiplying the energy term in Eq. (19) by a Boltzmann factor $e^{\varepsilon/2kT}$. A plot of $\langle \Delta n \rangle^2 / \bar{n}$ vs. ε/kT based on Eq. (40) is shown in Fig. 4.

3. High Coverage

By analogy to the arguments advanced for the simple lattice gas we should expect fluctuations to decrease at high particle coverage, corresponding to low hole coverage. It can be shown that

$$\frac{\langle \Delta n \rangle^2}{m} = \frac{\phi}{1 + (\varepsilon/kT)\phi g(\phi)} \quad (42)$$

where $\phi = x(1-x)$ and $g(\phi)$ is a function of ϕ . Thus the absolute mean square particle fluctuation is a symmetric function of mean coverage about $x = 0.5$, since the factor $x(1-x)$ is symmetric about this point. The arguments leading to Eq. (42) are tedious and will only be outlined here. We note from Eqs. (26) and (27) that \bar{y} can be written as

$$\bar{y} = \frac{zn}{2} (1-2(1-x)f(\phi)) \quad (43)$$

and that integration with respect to β leads to

$$y_o = \frac{zn}{2} (1-2(1-x)f_o(\phi)) \quad (44)$$

where f , and f_o are different functions of ϕ . Double differentiation of y_o with respect to m leads to

$$\frac{\partial^2 y_o}{\partial m^2} = \frac{zx^2}{m} g(\phi) \quad (45)$$

where $g(\phi)$ is a function of ϕ only, i.e., does not depend on x or $(1-x)$ separately, so that Eq. (42) results.

II. Two Coexisting Phases

So far we have considered only a single phase of the adsorbed layer. While it seems probable that the results on chemisorbed layers obtained to date by Chen and Gomer² can be explained in these terms, it is worthwhile to explore other situations as well, since these could occur for instance in physisorption, which may also be amenable to study by the field emission technique outlined at the beginning of this paper.

Specifically we consider here two adsorbed phases in equilibrium. This implies attractive interactions and a temperature below the (two-dimensional) critical T , T_c . We carry out the analysis on the basis of the simple lattice gas model, whose free energy is given by Eq. (19).

For equilibrium between two phases, $i = 1, 2$, we require that $\mu_1 = \mu_2$ and also that $p_1 = p_2$, where the chemical potential μ is given by

$$\mu = \left[\frac{\partial F}{\partial n_i} \right]_{T, m_i} \quad (46)$$

and the two-dimensional pressures p_s by Eq. (20). One obtains from Eqs. (19), (46) and (20)

$$2n\bar{x}_1 + \ln \frac{\bar{x}_1}{1-\bar{x}_1} = 2n\bar{x}_2 + \ln \frac{\bar{x}_2}{1-\bar{x}_2} \quad (47)$$

and

$$-n\bar{x}_1^2 + \ln(1-\bar{x}_1) = -n\bar{x}_2^2 + \ln(1-\bar{x}_2) \quad (48)$$

for the conditions of equilibrium in a closed system; all \bar{x}_i are average values, as usual. n is defined as before as $ze/2kT$. Equations (47) and (48) are also obtained by minimizing the total free energy for the two phase system as a function of n_1 and m_1 .

Equations (47) and (48) are satisfied by the trivial solution $\bar{x}_1 = \bar{x}_2$, but also by

$$\bar{x}_1 + \bar{x}_2 = 1 \quad (49)$$

which leads to

$$n(1-2\bar{x}_2) = \ln \frac{\bar{x}_2^2}{1-\bar{x}_2} \quad (50)$$

which can be solved numerically for \bar{x}_2 for given n or trivially for n , for given \bar{x}_2 . It is also possible to obtain the critical temperature, or simply the critical value of n , n_c from the requirement that

$$\frac{\partial p}{\partial m} = \frac{\partial^2 p}{\partial m^2} = 0 \quad (51)$$

at n_c . This leads straightforwardly to $n_c = 2$.

We consider now the fluctuations to be expected in an open subregion of the (infinite) closed system. We define the total number of sites in the open subregion as m , and the number of sites occupied by phase 1 in A as m_1 , those occupied by phase 2 as m_2 and the number of ad-particles in A belonging to phase 1 as n_1 , those belonging to phase 2 as n_2 , and define $x_1 = n_1/m_1$, $x_2 = n_2/m_2$. It is to be noted that

$$m_1 + m_2 = m \quad (52)$$

since the number of sites in A is fixed, but that n_1 , x_1 , n_2 , and x_2 obey no conservation rules, other than $x_i \leq 1$, $n_i \leq m$. However the average values, $\bar{x}_1 + \bar{x}_2 = 1$ as before. We now define the following quantities: Q_{m_1} is the probability of a partition into m_1 , $m-m_1$ sites for phases 1 and 2 respectively. Since there are m ways of dividing m sites into two groups,

$$Q_{m_1} = 1/m \quad (53)$$

$P^{m_1}(n_1)$ is the probability of there being n_1 molecules in phase 1 in A, if there are m_1 sites belonging to phase 1 in A. $P^{m-m_1}(n_2)$ is the probability of there being n_2 molecules in phase 2 in A if there are $m_2 = m-m_1$ sites belonging to phase 2 in A. It is to be noted that $P^{m_1}(n_1)$ and $P^{m-m_1}(n_2)$ are independent. Then $Q_{m_1} P^{m_1}(n_1) P^{m-m_1}(n_2)$ is the probability of a partition with m_1 sites assigned to phase 1, and n_1 and n_2 ad-particles in phases 1 and 2, respectively. Finally we define $\langle \Delta_{m_1} n_1 \rangle^2$ as the mean square fluctuation of n_1 under the partition m_1 , $m-m_1$, $\langle n_1 \rangle_{m_1}$ as the mean number of ad-particles in phase 1 in A, under the partition m_1 , $m-m_1$, and $\langle \Delta_{m-m_1} n_2 \rangle^2$ and $\langle n_2 \rangle_{m-m_1}$ as the corresponding quantities, under the same partition for n_2 .

We can then write for the total fluctuation in region A

$$\begin{aligned} \langle \Delta n \rangle^2 &= \sum_{m_1} \sum_{n_1} \sum_{n_2} Q_{m_1} P^{m_1}(n_1) P^{m-m_1}(n_2) (n_1 + n_2)^2 \\ &\quad - \left\{ \sum_{m_1} \sum_{n_1} \sum_{n_2} Q_{m_1} P^{m_1}(n_1) P^{m-m_1}(n_2) \right\}^2 \end{aligned} \quad (54)$$

Carrying out the summations over n_1 , n_2 and noting that

$$\sum_{n_1}^{m_1} P^{m_1}(n_1) (n_1^2) = \langle n_1^2 \rangle_{m_1} \quad (55)$$

and so forth, we have

$$\begin{aligned}
\langle \Delta n \rangle^2 &= \sum_{m_1} Q_{m_1} \langle n_1^2 \rangle_{m_1} + \sum_{m_1} Q_{m_1} \langle n_2^2 \rangle_{m_1} + 2 \sum_{m_1} Q_{m_1} \langle n_1 \rangle_{m_1} \langle n_2 \rangle_{m_1} \\
&- \left(\sum_{m_1} Q_{m_1} \langle n_1 \rangle_{m_1} \right)^2 - \left(\sum_{m_1} Q_{m_1} \langle n_2 \rangle_{m_1} \right)^2 \\
&- 2 \left(\sum_{m_1} Q_{m_1} \langle n_1 \rangle_{m_1} \right) \left(\sum_{m_1} Q_{m_1} \langle n_2 \rangle_{m_1} \right) \tag{56}
\end{aligned}$$

Since

$$\langle n_1^2 \rangle_{m_1} = \langle \Delta_{m_1} n_1 \rangle^2 + \langle n_1 \rangle_{m_1}^2 \tag{57}$$

with an analogous expression for $\langle n_2^2 \rangle_{m_1}$ we obtain

$$\begin{aligned}
\langle \Delta n \rangle^2 &= \sum_{m_1} Q_{m_1} \langle \Delta_{m_1} n_1 \rangle^2 + \sum_{m_1} Q_{m_1} \langle \Delta_{m_1} n_2 \rangle^2 + \sum_{m_1} Q_{m_1} \langle n_1 \rangle_{m_1}^2 \\
&+ \sum_{m_1} Q_{m_1} \langle n_2 \rangle_{m_1}^2 + 2 \sum_{m_1} Q_{m_1} \langle n_1 \rangle_{m_1} \langle n_2 \rangle_{m_1} - \left(\sum_{m_1} Q_{m_1} \langle n_1 \rangle_{m_1} \right)^2 \\
&- \left(\sum_{m_1} Q_{m_1} \langle n_2 \rangle_{m_1} \right)^2 - 2 \left(\sum_{m_1} Q_{m_1} \langle n_1 \rangle_{m_1} \right) \left(\sum_{m_1} Q_{m_1} \langle n_2 \rangle_{m_1} \right) \tag{58}
\end{aligned}$$

We note that

$$\langle \Delta_{m_1} n_1 \rangle^2 = \frac{kT K_1 \bar{x}_1^2}{m_1 a} = \left(\frac{kT}{a} \right) K_1 \bar{x}_1^2 m_1 \tag{59}$$

and

$$\langle \Delta_{m_1} n_2 \rangle^2 = \left(\frac{kT}{a} \right) K_2 \bar{x}_2^2 (m - m_1) \tag{60}$$

Replacing \sum_{m_1} by $\int_0^m dm_1$ we finally obtain

$$\langle \Delta n \rangle^2 = \frac{mkT}{2a} (\bar{x}_1^2 K_1 + \bar{x}_2^2 K_2) + \frac{m^2}{12} (\bar{x}_2 - \bar{x}_1)^2 \tag{61}$$

The first term in Eq. (61) represents the contribution from fluctuations within the two phases. It is interesting that on average our model predicts half the sites in A to be covered by each phase which accounts for the factor of 2 in the denominator of the first term. The second term in Eq. (61) corresponds to fluctuations in the partition between phases, and is proportional to the square of the number of sites in A, or to A^2 .

K_1 and K_2 can be evaluated explicitly by means of Eq. (20) and lead to the result, already familiar from Eq. (21)

$$\langle \Delta n \rangle^2 = m \frac{\bar{x}_1 (1-\bar{x}_1)}{2\eta \bar{x}_1 (1-\bar{x})+1} + \frac{m^2}{12} (\bar{x}_1 - \bar{x}_2)^2 \quad (62)$$

Fluctuations in the Dilute Phase Only

Up to this point we have assumed the open subregion to contain both dilute and dense phases and have calculated the total fluctuation in number density. It is easy, of course, to calculate also the fluctuations in the dilute phase alone, when the region of observation contains both phases and when it contains only the dilute phase. The latter situation might conceivably occur if the dense phase is anchored to the edge of a plane of the field emitter. The results are given by Eq. (21) with the following modification: If the region of observation contains both dense and dilute phases, the dilute phase occupies, on average, half of the region and hence $\langle \Delta n \rangle^2$ dilute is given by the result of Eq. (21) divided by 2. If the observation

region contains only the dilute phase, Eq. (21) applies without modification.

Phase 2 Always Dense

The analysis given here has omitted line tension, or better line energy effects, i.e., the two-dimensional analogue of surface energy effects. These are twofold. First, they will tend to prevent the segregation of the dense phase into many small patches. Second, they will have the effect of keeping the density of the dense phase high. We simulate this by postulating that $\bar{x}_2 = 1$ at all T. An analysis analogous to that leading to Eq. (50) then shows that

$$n(1-2\bar{x}_1) = \ln \frac{\bar{x}_1}{1-\bar{x}_1} \quad (63)$$

which is identical in form to Eq. (50). However, $\bar{x}_1 + \bar{x}_2 \neq 1$ here. Proceeding by steps analogous to those leading from Eq. (54) to Eq. (62) we find that $\langle \Delta n \rangle^2$ is identical to that obtained by letting $\bar{x}_2 = 1$ in Eq. (62). Since the solution of Eq. (62) for \bar{x}_1 is identical to that obtained for \bar{x}_2 from Eq. (50) and since for that case $\bar{x}_1(1-\bar{x}_1) = \bar{x}_2(1-\bar{x}_2)$ we need not make any change in Eq. (62) other than to set $\bar{x}_2 = 1$ in the second term. It further follows that the fluctuations in phase 1 alone are still given by both the form and numerical values, for given x_1 , of Eq. (21), if only the dilute phase exists in the region of observation, or by half that value if both phases occur in it.

Table II shows values of $\bar{x}_1(1-\bar{x}_1)/(1+2\eta\bar{x}_1(1-\bar{x}_1))$, $(m/12)(\bar{x}_1-\bar{x}_2)^2$ and $\langle\Delta n\rangle^2/m$ as functions of η according to the unrestricted form of Eq. (62). These data are also displayed in Fig. 5. Several interesting conclusions can be drawn from these results. First, we note that far below the critical temperature ($|\eta| \gg 1$) fluctuations in the partition between the phases, i.e., the second term in Eq. (62) dominate and are effectively constant since $\bar{x}_2 \gg \bar{x}_1$. As the temperature rises, $\bar{x}_2 \rightarrow \bar{x}_1$ and this term decreases. The fluctuations thus decrease with increasing temperature. This is, of course, a result we might have anticipated since the two phase system is in fact a special case of attractive interactions. However, very near the critical temperature, i.e., as $|\eta| \rightarrow 2$, fluctuations within each phase become very large and dominate the total fluctuations. Thus the curves of Fig. 4 go through a minimum and then rise steeply. To date this behavior has not in fact been observed but should be expected for physisorption systems near T_c .

III. Equilibrium Between Singlets and Doublets

We conclude by discussing very briefly a special case of attractive interactions. Suppose that a given adsorbate B can exist as single ad-particles B and also as doublets B_2 , and suppose that only the single particles B are mobile. For this argument B_2 need not be an adsorbed molecule; it suffices that there be an attractive interaction between two B particles, and that an adjacent filled pair of sites constitute an immobile

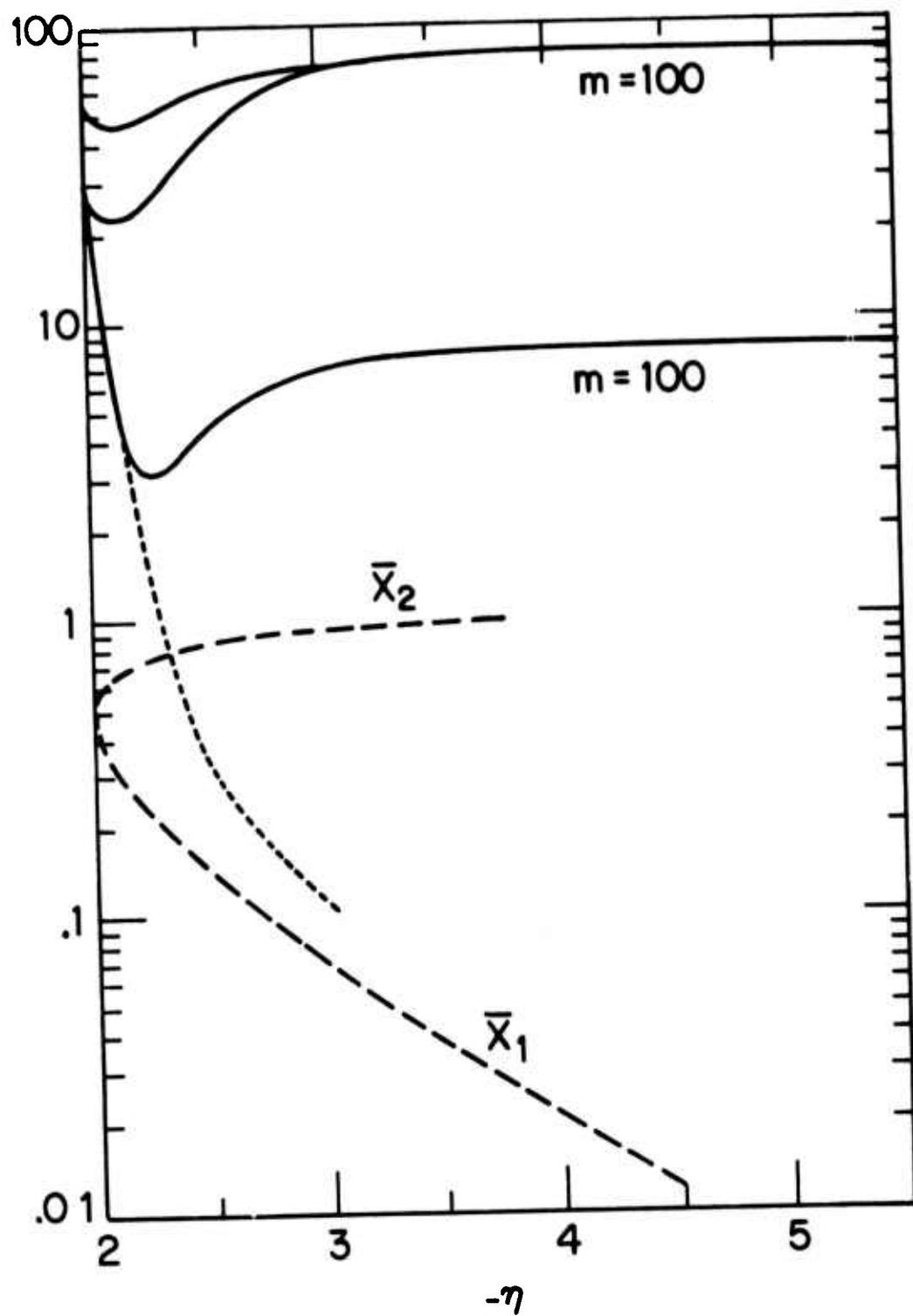


Figure 5. Various quantities relevant to the two-phase situation plotted vs. η : Solid lines: $\langle \Delta n \rangle^2 / m$ for $m = 100, 1000$, plotted according to the unrestricted case of Eq. (62); for $m = 1000$ the curve corresponding to $\bar{x}_2 = 1$ is also shown; dashed lines: \bar{x}_1 , \bar{x}_2 ; dotted line: $\bar{x}_1(1-\bar{x}_1)/[1+2\eta(\bar{x}_1(1-\bar{x}_1))]$.

entity. It might be thought at first that the fluctuations in such a system will increase with increasing T , since the number of mobile particles, B , increases with increasing T . However, the observed fluctuations refer to both B and B_2 , and fluctuations in the latter are caused via the equilibrium



with B mobile. A straightforward analysis, based for simplicity on the classical form of the free energy

$$F_i = n_i kT \ln x_i + \epsilon n_i - kT n_i \quad (64)$$

and equilibrium between B and B_2 with an energy of dissociation ϵ leads to the result

$$\frac{\langle \Delta n \rangle^2}{\bar{n}} = \frac{(1+4\bar{x}e^{-\epsilon/kT})^{\frac{1}{2}}}{1+(1+4\bar{x}e^{-\epsilon/kT})^{\frac{1}{2}}} \quad (65)$$

where \bar{x} is the mean total adsorbate concentration in terms of B . Thus $\langle \Delta n \rangle^2$ decreases as T increases.

Conclusion

The foregoing indicates that a study of mean square concentration fluctuations in small open domains of adsorbed layers can provide valuable information on the thermodynamics and statistical mechanics of such layers, including the magnitude and sign of ad-particle interactions, and the occurrence of phase transitions and critical points.

Acknowledgements

We wish to thank Professor Brice Carnahan of the Department of Chemical Engineering of The University of Michigan for carrying out the computer calculations used in this work. This research was supported in part by NSF Grant, NSF DMR 73-07589-A01 and by the Advanced Research Projects Agency of the Department of Defense under Contract DAHC15-71-C-0253 with The University of Michigan.

References

1. R. Gomer, Surface Science 38, 373 (1973).
2. J. R. Chen and R. Gomer, to be published.
3. N. Davidson, Statistical Mechanics, McGraw-Hill, New York (1962).
4. F. Reif, Fundamentals of Statistical and Thermal Physics, McGraw-Hill, New York (1965).
5. T. L. Einstein and J. R. Schrieffer, Phys. Rev. B7, 3629 (1973).
6. L. D. Schmidt and R. Gomer, J. Chem. Phys. 45, 1605 (1966).
7. R. Fowler and E. A. Guggenheim, Statistical Thermodynamics, pp. 429-441, Cambridge University Press, Cambridge, England (1952).

TABLE I. Values of $I' = \int_1^{\infty} (e^{-\alpha x^3} - 1) x dx$

α	I'
0.1	-9.875×10^{-2}
0.5	-4.714×10^{-1}
1.0	-8.950
5.0	-3.416
7.0	-4.400
10.0	-5.715
20.0	-9.365

$x = r/r_o$; r_o is minimum separation of 2 ad-particles.

$\alpha = u_o/kT$; u_o is (repulsive) interaction energy at r_o .

TABLE II. Fluctuations in a 2 Phase Ad-Layer

$-\eta$	$\frac{x(1-x)}{1+2nx(1-x)}$	$(x_1 - x_2)^2 / 12$	for $m = 100$	$\langle \Delta n \rangle^2 / m$ for $m = 1000$
2.0005	499	6×10^{-5}	499	499
2.01	24.8	1.2×10^{-3}	24.9	26
2.5	0.33	0.042	4.52	42.3
3.0	0.11	0.071	7.19	70.9
5.0	0.007	0.081	8.10	80.9
10.0	4.5×10^{-5}	0.083	8.33	83.3
16.0	10^{-7}	0.083	8.33	83.3

FEASIBILITY OF OPERATING OF ENERGY CONVERSION
MACHINES UNDER REDUCING CONDITIONS AT VERY HIGH TEMPERATURES

E. E. Hucke

Discussion with various members of the MRC have been held concerning the idea of a very high temperature energy conversion machine. The basic chemistry of any practical (non-fuel cell) combustion of fossil fuel require that the reaction products have a high activity of carbon, hydrogen and nitrogen, but there is no fundamental reason that the flame have a high oxygen potential. Flames at stoichiometric conditions or slightly fuel rich attain almost the same flame temperature as highly oxidizing flames. In light of this, the inherent properties of the various types of carbon materials have been explored. Such materials have either equal or vastly superior mechanical and thermal properties to any others known. Oxidation resistance is the one limiting feature. In reducing conditions one or more of the carbon materials, which are non-strategic, relatively cheap and formable can easily operate at high stress levels at temperatures exceeding 2000°C. The adiabatic flame temperatures of most fossil fuels and their derivatives is approximately 2000°C. A preliminary discussion session involving members of the Council and representatives from Livermore Laboratories and EPRI was organized to further analyze the problem areas likely to be encountered in such a very high temperature machine.

ACID RAINWATER

J. L. Margrave

ABSTRACT

Weather reports which recorded the pH of rainwater are scarce, but in the limited number of cases for which data are available, there is a large apparent trend toward increasing acidity in rainwater. For example, pH~6 in the 1930's but pH ~4-5 now in New Hampshire/Upper New York State. The bio-effects are unknown. Weathering of rocks, soil and buildings will be accelerated. SO_2/SO_3 from fossil fuels react to yield H_2SO_4 ; there is some HNO_3 , too. This problem could get worse as coal/lignite/shale based fuels are used more extensively for energy generation, unless sulfur is removed.

Among the potential solutions to this problem considered are (a) The complete removal of sulfur from fuels; (b) The use of neutralizing additives to form sulfates which can be precipitated; (c) The use of photo-excitation (UV lamps/lasers) to activate SO_2 , and accelerate oxidation to SO_3 and nucleation of H_2SO_4 before release of stack gases.

ACID RAINWATER

J. L. Margrave

I. THE PROBLEM - HOW TO CONTROL THE ACIDITY OF RAINWATER

The chemistry of precipitation falling on various parts of the United States, England and Europe has been observed and evaluated by a number of investigations over the past 50 years. Data are scattered and incomplete but there is a pattern of increasing acidity of rainfall, which is now building into an increasing acidity of the lakes into which this rain drains.

Such acid rain causes leaching of calcium and other nutrients from the soil and from plant foliage, growth rates of forests are reduced, seedlings appear more susceptible to disease, and serious fish mortality has been reported. Damage to buildings, roadways, statues, etc. could also be extensive.

II. OBSERVED TRENDS AND EFFECTS

A. Before 1930

According to Likens et al.,¹ the records of acidity data for this period are rare, and mainly qualitative. For example, samples collected prior to 1930 contained relatively large amounts of HCO_3^- , indicative of $\text{pH} \geq 5.6$, and methyl orange indicator tests all showed $\text{pH} > 4.6$ in early water measurements. The ionic contaminants were mainly $\text{SO}_4^{=}$, with

amounts of NH_4^+ and NO_3^- . The main source of $\text{SO}_4^=$ was from the $\text{SO}_2/\text{SO}_3 \rightarrow \text{H}_2\text{SO}_4$ reaction from the sulfur contained in coal being burned for heat/electrical generation purposes.

B. 1930 - 1955

During the period 1930-1955 there were four major factors which one would have expected to affect the acidity of rain:

(1) Less coal was burned and many plants converted to natural gas. Gas heating for houses became more prevalent.

(2) The chimneys which exhausted combustion products were higher and higher, injecting SO_2 into the atmosphere at heights of 60 to 360 meters.

(3) Mechanical and electrostatic precipitations were removing at least 50%, and often more than 75% of the particulate matter - fly ash.

(4) The use of automobiles increased greatly.

The results of these presumed technological advances were, paradoxically, to create a regional trend toward lower $\text{SO}_4^=$, more NO_3^- and higher pH in the precipitation. Thus, in England, Gorham³ reported the pH of rains in the lake district during the summer of 1954 to fall in the 4.0-5.8 range with the average pH ~4.45, the general association of this acidity with $\text{SO}_4^=$ (and H_2SO_4) was clearly established. Tamm⁴ gave a pH range of 4.3-5.9 for rains in open fields of East Sweden in 1953, and decreasing pH's of lakes and rivers are attributed to acid rains.⁵

C. 1955 - 1975

Data for this period are extensive and include ionic concentrations, pH, and soot concentrations along with worldwide meteorological observations. Likens and Bormann¹ report NO_3^- to increase by a factor of 3, $\text{SO}_4^=$ to decrease slightly and pH to vary over the range 4.0-4.20 for the 1964-1972 period.

The neutralizing character of fly ash, now being ~99% eliminated from chimney emissions, was quantitatively established by Shannon and Fine⁶, and the unexperiencec continual increase in acidity in spite of lowered $\text{SO}_4^=$ clearly shown. Soot particles in the emission of chimneys apparently played a useful role by serving as nuclei for SO_3 adsorption, H_2SO_4 formation and as a partially neutralizing environment before the advent of highly efficient precipitator systems. Newman² argues against this view and points out the statistical odds against SO_3 being really neutralized on a mole for mole basis by the fly ash.

During 1965-1970, increasing pH-trends for lakes/rivers in Sweden⁵, for lakes in Canada west of the Sudbury smelters⁷ and for lakes in the Adirondacks⁸ have been reported with values down to 3.9. Gorham² was concerned with the effect of such acidity on rock/soil weathering and observed that "...pitting of building stone in industrial cities is well known...", also from the action of acid rainwater.

III. SUMMARY AND SUGGESTIONS FOR FUTURE WORK

In reviewing this important topic, one of the clearest conclusions is that not enough data, even on a regional basis in the United States, are available to permit a clear statement of facts. Hopefully, both meteorologists and environmental scientists will be stimulated to begin extensive collection of samples, with follow-up analyses, to provide a more reliable data base, including pH and ionic concentrations.

The real-world approach in the 1930-1960 period to the SO_3 /fly ash emission problem, which is not almost universal, was to collect the fly ash and soot particles and to emit SO_2 to the atmosphere from very tall chimneys. The neutralizing effect of the fly ash on the acid-forming SO_2/SO_3 effluent has been lost and a man-made chemical factor of major proportions is now active in rainwater.

The simple answers to eliminate the problem are:

1. Reduce sulfur contents of fuels
2. Reduce NO_x emissions from natural gas burners and from automobiles.
3. Minimize industrial emissions of HCl , HF , SiCl_4 , BCl_3 , and other compounds which are acidic or hydrolyze to produce acidic solutions.

The achievement of these goals is expensive and somewhat inconsistent with economic pressures to burn more coal/lignite/shale fuels and solve energy problems, to soften enforcement of SO_3 and soot emission requirements and to soften

automobile emission standards to conserve gasoline.

One area of research is clearly of great importance, the study of the bio-effects of acid rain, food plants, forests and our grass supply for grazing animals are crucial resources and man must use great caution when he creates world-wide atmosphere conditions that could interfere with these natural processes.

Another type of question - can one find an additive that would up the SO_2/SO_3 , in the form of a volatile sulphate species that would not hydrolyze appreciably? In principle, something like Na_2CO_3 , or K_2CO_3 , to react with SO_2/SO_3 , should yield Na_2SO_4 , which is volatile with only partial decomposition or K_2SO_4 , which is even more stable as a molecular gaseous species. The high temperature behaviors of MgSO_4 , CaSO_4 , etc., are currently uncertain but it would be interesting to know if one can carry amounts of gaseous metal oxides or gaseous sulfates through electrostatic precipitations and whether or not such additives would affect $\text{SO}_2/\text{SO}_3/\text{H}_2\text{SO}_4$ concentrations.

Of course, the original ash content of lignites and many coals really supplies such an additive, and the study of Shannon and Fine⁶ reports that approximately 50% of the soluble cations in fly ash from dust collectors appear to be present as sulfates. The electrostatic precipitator ash contains less sulfate. The identification of an additive to capture SO_3 , and form stable sulfates appears likely and could solve this problem.

Another approach to the problem is to activate the effluent SO₂, probably with an intense UV source, to attain a reactive excited state which would oxidize and/or nucleate on inorganic ions or OH or H₂O molecules⁹.

ACKNOWLEDGEMENT

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

References

1. G. E. Likens and F. H. Bormann, *Science* 184, 1176 (1974);
ibid, 188, 958 (1975).
2. L. Newman, *Science* 188, 957 (1975).
3. E. Gorham, *Geochim. Cosmochim. Acta* 7, 231 (1975).
4. C. O. Tamm, *Medd. Skogsforskn Inst. Stockholm* 43, 1 (1953).
5. S. Oden and T. Ahl, *Sartryck UR Ymer Arsbok*, 1970, pp. 103-122.
6. D. G. Shannon and L. O. Fine, *Env. Sci. Tech.* 8, 1026 (1974).
7. (a) R. J. Beamish and H. H. Harvey, *J. Fish Res. Board Can.* 29, (8), 1131 (1972);
(b) R. J. Beamish, *Water Res.* 8, 85 (1974).
8. C. L. Schofield, Jr., *Trans. Am. Fish Soc.* 94, (3), 227 (1965).
9. R. Heist and H. Reiss, *J. Chem. Phys.* 59, 665 (1973);
ibid, G1, 573 (1974).

AN APPROACH FOR PREPARING ANTIFOULING
CONCRETE FOR UNDERSEA USE

C. U. Pittman, Jr.

ABSTRACT

Currently, antifouling concrete is made by impregnating porous shale with creasote oil and tin compounds followed by mixing with portland cement. It is fairly weak and is used for compressive loads up to 3000 psi and uses large amounts of biocide. Polymer impregnated concrete (PIC) made by impregnating mature concrete with a liquid monomer, which is then polymerized in-situ, develops compressive strengths greater than 20,000 psi and shows good durability against chemical attack and excellent impermeability. It is proposed that liquid monomers, which contain functions biocidal to the fouling community, be mixed with the monomers used in the polymer impregnating process. Then a PIC which is antifouling should result having high compressive strengths. Similarly, the use of biocidal monomers in the area of polymer strengthened concrete (PSC) should also be investigated.

AN APPROACH FOR PREPARING ANTIPOULING CONCRETE
FOR UNDERSEA USE

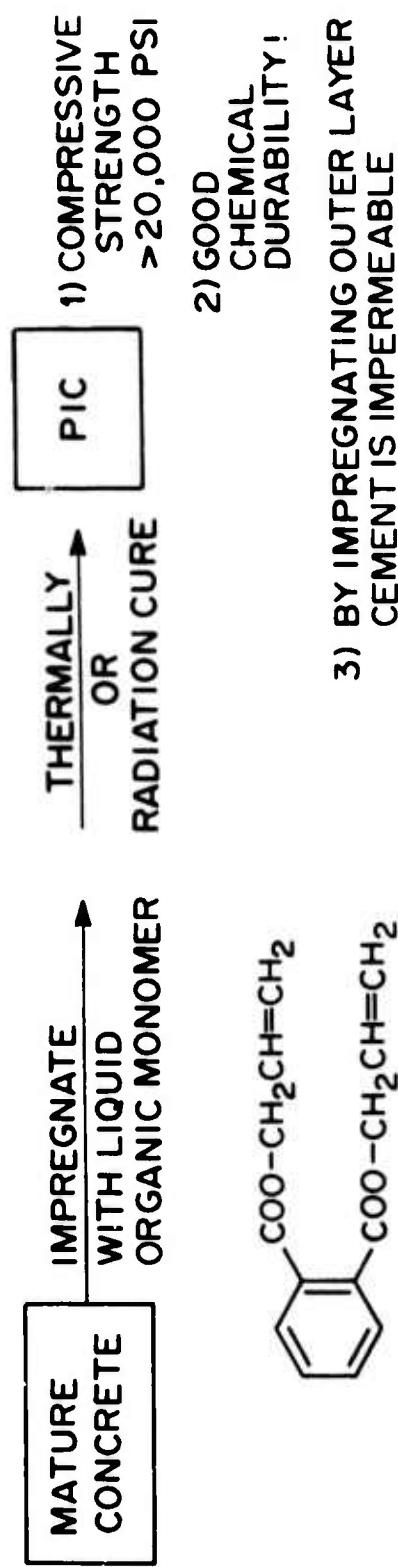
C. U. Pittman, Jr.

I. INTRODUCTION

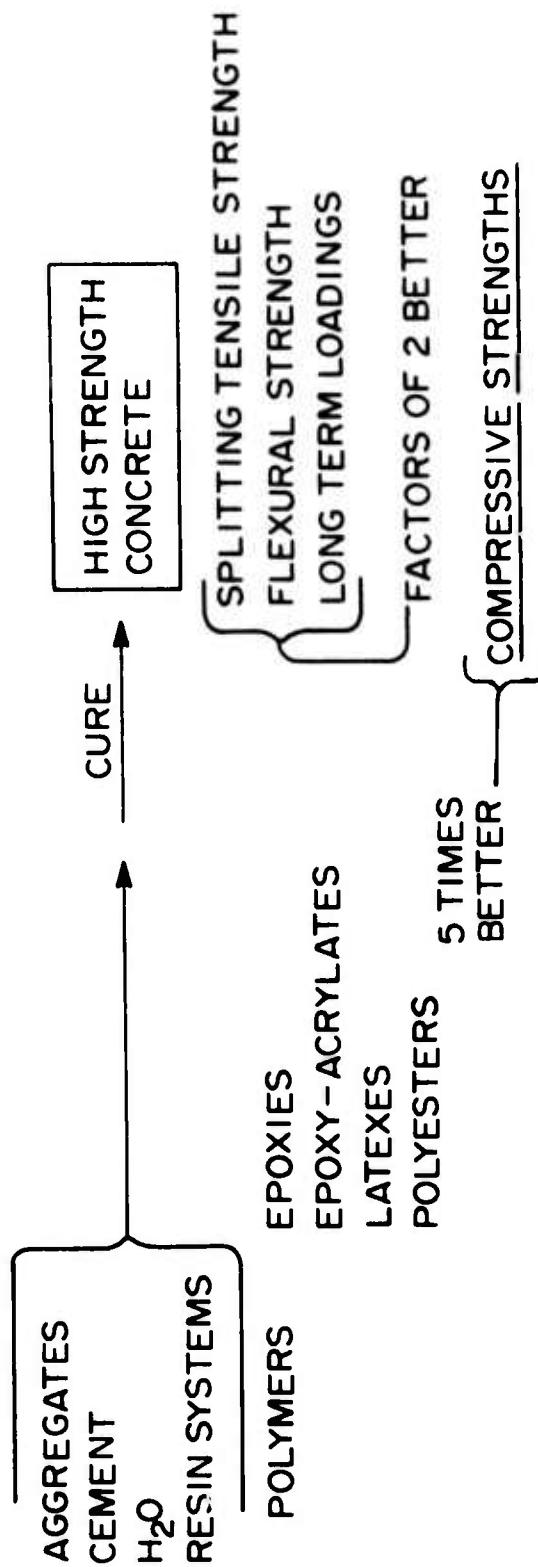
Polymer-Impregnated Concrete (PIC) is made by impregnating mature concrete with a liquid monomer, which is subsequently polymerized in-situ, either thermally or via a radiation curing process. The resulting material contains 6% to 8% by weight of polymer, in general. Such concretes develop compressive strengths over 20,000 psi and show good durability against chemical attack and excellent impermeability. These properties are ideally suited to underwater pressure-resistant concrete structures. This is outlined in Figure 1.

The Civil Engineering Laboratory, Ocean Structures Division, Port Hueneme, California, had prepared and tested 16 inch OD diameter spheres of PIC where the polymer was a polydiallylphthalate (8.4% by weight of the concrete). The spheres exhibited linear elastic strain behavior and implosion pressures were predictable using classical elastic theory. Spheres, where only the outer portion of the concrete wall was impregnated (to 1 in depth), were also made to prevent impregnation by sea water. PIC resisted water permeation into the spheres at submersion depths between 2000 and 5000 ft.

FIGURE 1. POLYMER IMPREGNATED CONCRETE (PIC)



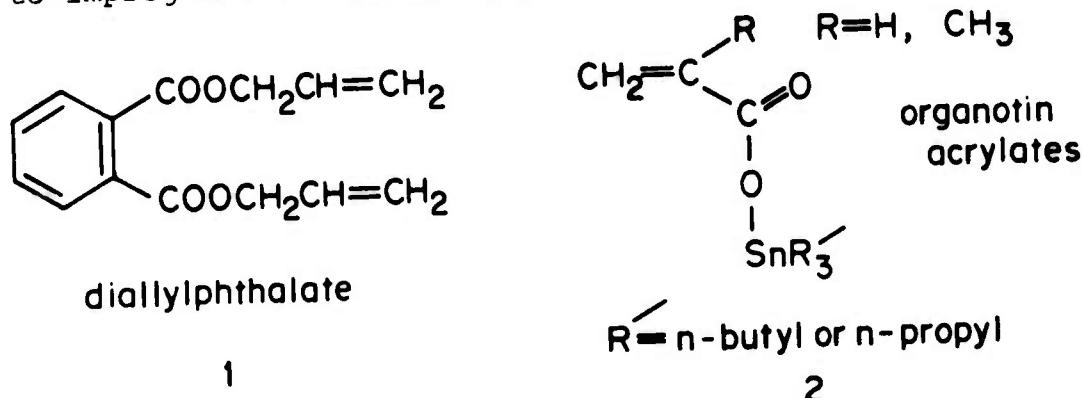
POLYMER STRENGTHENED CONCRETE (PSC)



However, PIC which is to be used for long term underwater tasks, will come under attack by fouling organisms which could pit the surface or otherwise impede their use. Thus it would appear that the impregnating polymer should be one that contains a chemically anchored biocide. Recently, trialkyltin esters have been investigated in antifouling coatings. They have successfully been used to reduce the rate of biocide leaching (versus simple addition of the biocide to the coating) while effectively preventing development of the fouling community. Therefore, organotin monomers would be an ideal starting point for making polymer impregnated biocidal concrete CPIBC).

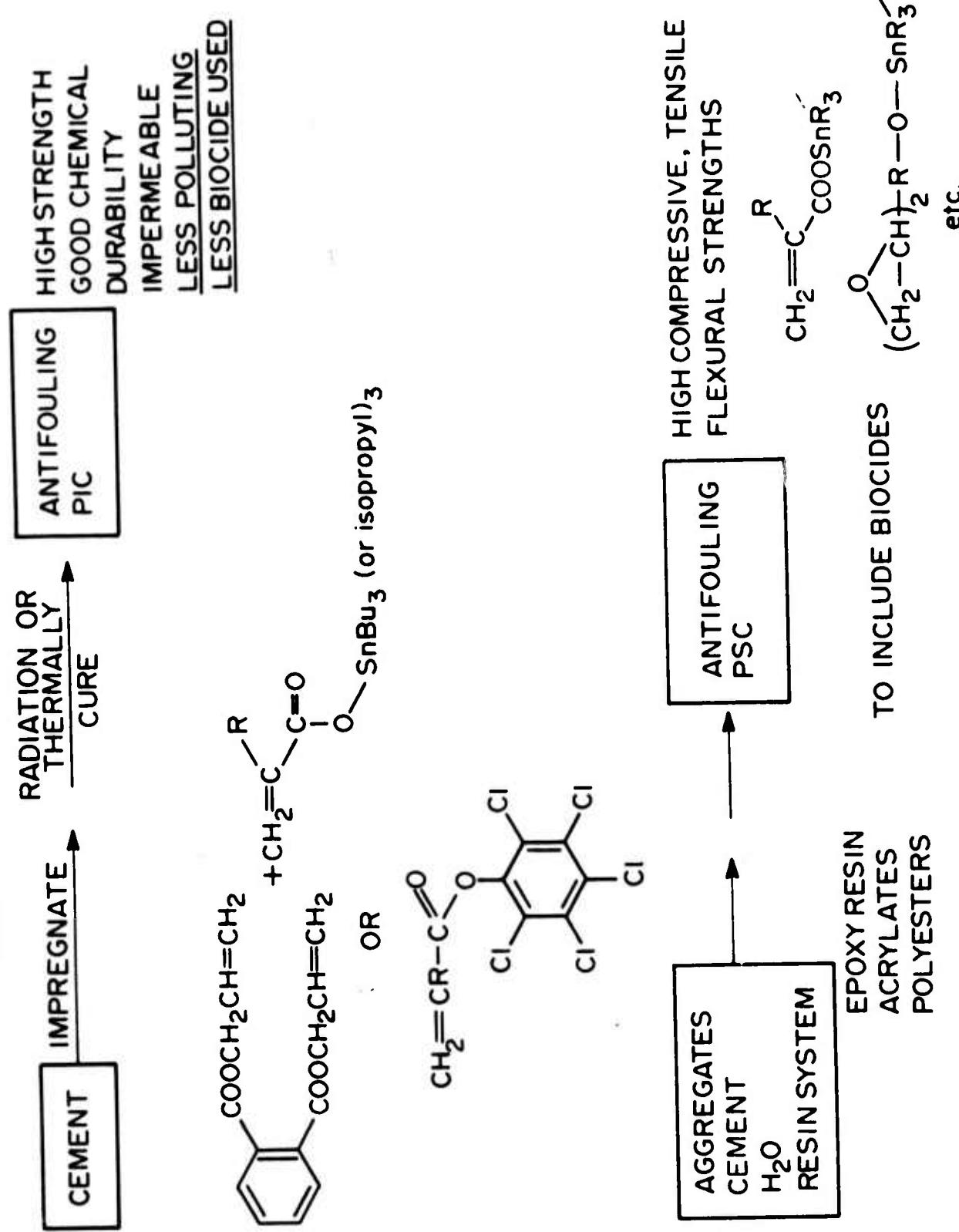
II. PROPOSED STUDIES

Diallylphthalate and organotin acrylates could be used to impregnate concrete, followed by curing (see Figure 2).

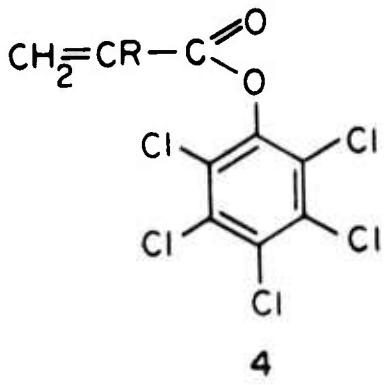
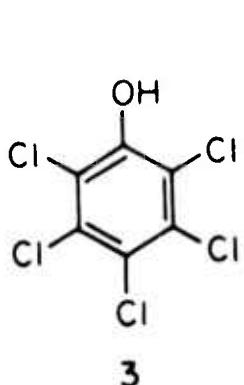


In this process, the organotin acrylate would be incorporated into the polymer via its vinyl function. The biocide need not be limited to organotins. A variety of other materials are available which could be made into appropriate monomers

FIGURE 2. PROPOSED HIGH STRENGTH ANTIFOULING CONCRETE



for use in the process. For example, chlorinated phenols are toxic and compounds 4 and 5 might be suitable monomers which would release biocide 3 under conditions of biological attack.



Since a succession of organisms is involved in the fouling community, more than one broad spectrum biocide with proven antifouling capability could be used in the impregnation-polymerization process. It should be noted that this type of biocidal impregnation treatment need not be of the type which permeates the entire concrete structure. Instead, it could be limited to the outer layer (1 to 6 inch depth) of the concrete.

Currently, antifouling concretes are made by an entirely different process which results in a product with a lower strength than that envisioned here.

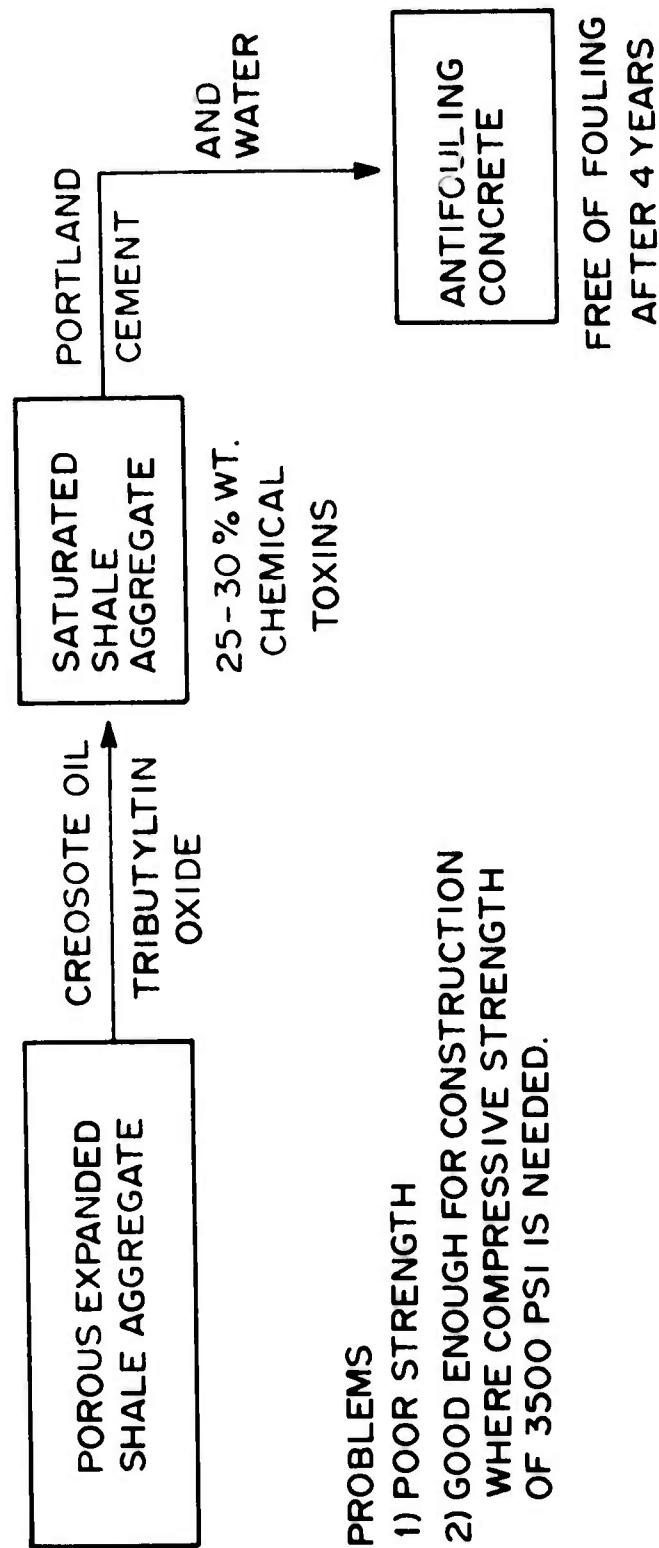
Naval research at the Civil Engineering Laboratory, OSD, Port Hueneme, California, has developed methods to incorporate toxic chemicals into concrete to protect it from the 2000 species of plants and animals that have been identified, to date, in the fouling community. The technique now under study at that location consists of impregnating a porous

expanded shale aggregate with mixtures of creosote oil and other toxic chemicals first. The porous, expanded, impregnated shale aggregate has absorbed the chemical solution to the extent of 25-30% of its dry weight with the toxic chemicals constituting 10% to 12% of the weight of the concrete made, ultimately, with these aggregates. After the shale aggregate is initially impregnated, it is combined with portland cement and water. Figure 3 summarizes this process.

When creosote oil and trivinyltin oxide, at levels of 100 to 250 ml/l were used, the concrete remained essentially free of marine fouling for four years. Copper naphthenate proved to be useless as a biocide in these tests. The major problem was that this antifouling concrete was not as strong as concrete made from sand and gravel, although it was sufficiently strong for construction in which a compressive strength of 3500 psi was acceptable.

It appears that research to merge the polymer impregnation method with the current attempts to make an antifouling concrete could be of value. If one could polymer impregnate high strength concrete with biocidal polymers, one should obtain an antifouling concrete which a) is stronger, b) uses less biocide, c) leaches less biocide thereby reducing environmental damage to the surrounding seawater ecosystem.

FIGURE 3. ANTIFOULING CONCRETE
Made at the CEL of Port Hueneme, California



ACKNOWLEDGEMENT

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15=71-C-0253 with The University of Michigan.

A QUALITATIVE APPROACH TO ANISOTROPIC
SEMICONDUCTING POLYMERIC MATERIALS

C. U. Pittman, Jr.

ABSTRACT

Anisotropic electrical properties are needed for a variety of electrical devices. Block copolymers of a semiconducting (or conducting monomer), A, and an insulating monomer, B, could be prepared where phase separation in a regular array occurs to generate cylinders of A-domains in a matrix of B. This material should be semiconducting along the direction of the cylinder axis but insulating perpendicular to this axis. The longest cylinders which could be expected would be 10 μ . Similarly, interphase polymerizations can result in A-domains dispersed in a B matrix. In these polymers the A-domains could be much larger than those in block copolymers. This material could be cut or rolled into films or sheets which could be used as a one-dimensional semiconductor. Thus, the bulk film is a one-dimensional semiconductor, while the polymer which constitutes the A-domains may be a three-dimensional semiconductor. The possible synthesis of organic semiconductors from complexes found in "organic metals" is commented on. Three reports currently exist in the literature. Can improved semiconducting polymers be prepared using mixed valence complexes currently being studied for their metallic properties?

A QUALITATIVE APPROACH TO ANISOTROPIC
SEMICONDUCTING POLYMERIC MATERIALS

C. U. Pittman, Jr.

Anisotropic optical and electrical properties are needed for a number of electronic components. For example, anisotropic conductors are needed in display devices where an organic conducting material would be sandwiched between an EC film and a cubic device.¹ High strength and high conductivity are needed in this device, and it was proposed that $(SN)_x$ could be deposited where the molecular chain axis would be perpendicular to the plane of the EC film. The Penn group already has demonstrated that epitaxial films of $(SN)_x$ can be made on polyethylene and teflon films, but the deposition of epitaxial films on metals had not been achieved.² Further research in this area is needed.

A new approach for producing an anisotropically semiconducting (or in principal conducting) polymeric material is suggested here. This approach permits a three dimensional semiconductor (or conductor) to become a functional one-dimensional semiconductor. Consider a block copolymer³ composed of a semiconductor, A, and another monomer, B, which will impart suitable physical properties as well as insulating properties. In this concept, the A-block can be a three-

dimensional semiconductor (or conductor). However, in a block copolymer, like domains have a propensity to line up in ordered arrays. Styrene-butadiene block copolymers have been well studied in this respect. If a thin film of the B-A-B triblock copolymer was made and if the A-block to B-block size ratio was appropriately designed, domains of A blocks might permit a conducting path through the film, but not throughout the film. This concept is schematically (and perhaps hopefully) portrayed below. While this proposal will concentrate on block copolymers, analogous dispersed phase and interpolymerization processes exist to make analogous materials.³ The arrows represent the direction of current flow.

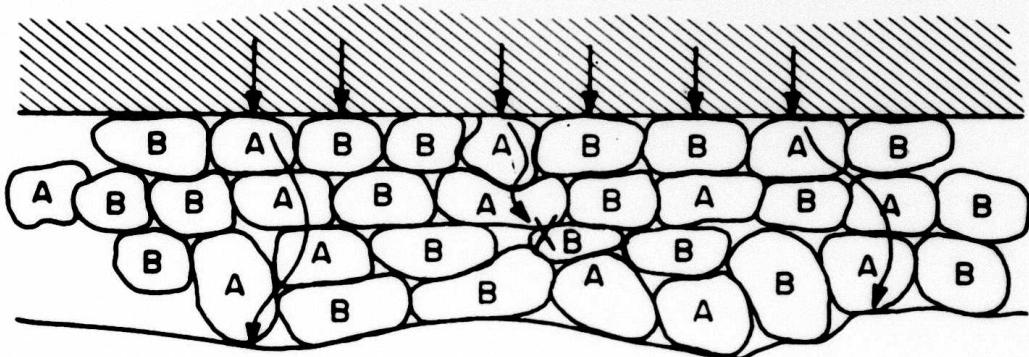


Figure 1.

The B-blocks are insulating, thus they prevent transmission of current while on the other hand, A-blocks will conduct. Where the A-blocks form a continuous path, the current would pass through the film. It is easily imagined that if block sizes are kept constant, the film will have

fewer conducting paths per unit surface, as the film gets thicker. This would occur if the A and B blocks have some disorder in their arrangement. It is also possible to imagine image blurring due to the hypothetical block morphology shown in Fig. 2.

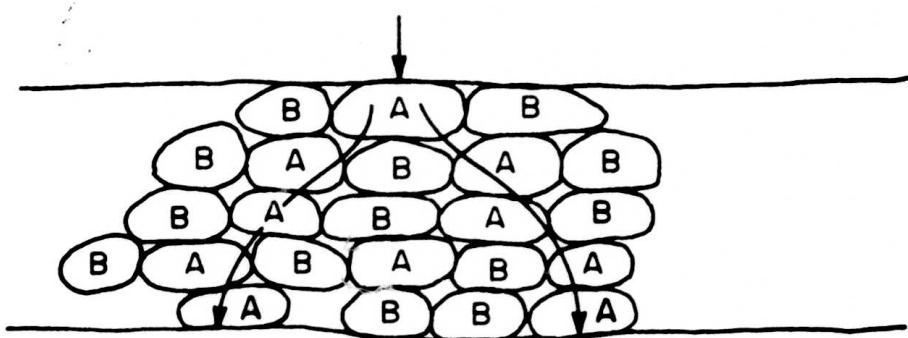


Figure 2.

Furthermore, the total conductivity, per unit cross sectional area would be lower in a block copolymer since the conducting path only constitutes a fraction of the surface. However, by varying total block size, relative block size (i.e., A vs. B), and film thickness, several parameters are available for optimization.

Another problem within this simple framework is the possible resistance to current flow which might exist at the interface between the A-blocks. For a semiconducting organic polymer similar interfaces exist in the homopolymer. Thus, the A-block domains should be similar to the homopolymer in their electrical behavior. However, for organic conductors

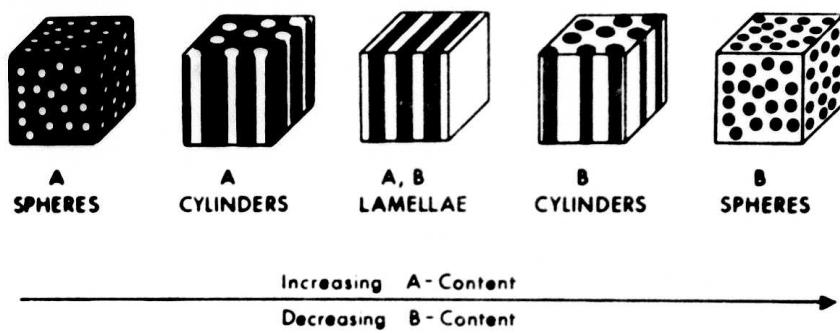


Figure 3. Schematic representation of the dependence of block copolymer morphology on the volume fraction of the blocks. A and B denote the two chemical species forming the molecule without reference to a particular system.

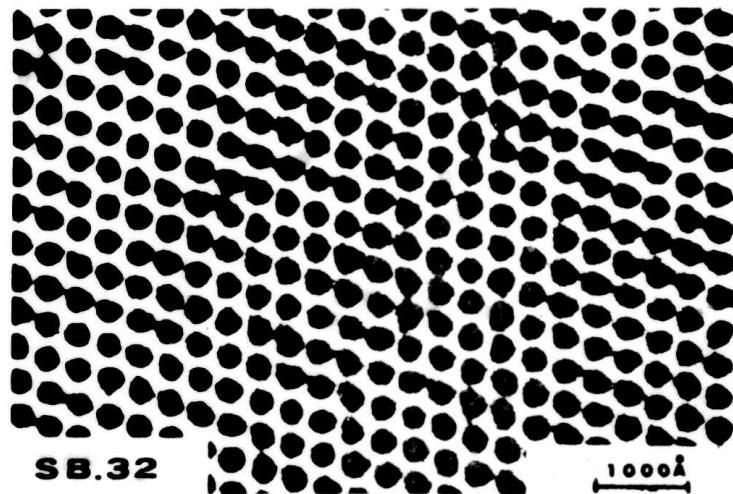


Figure 4. Electron micrograph of an S-B copolymer containing 28.3% S component. Here, as in all subsequent electron micrographs, the white regions correspond to the S phase while the black regions correspond to the osmium stained rubbery matrix.⁵

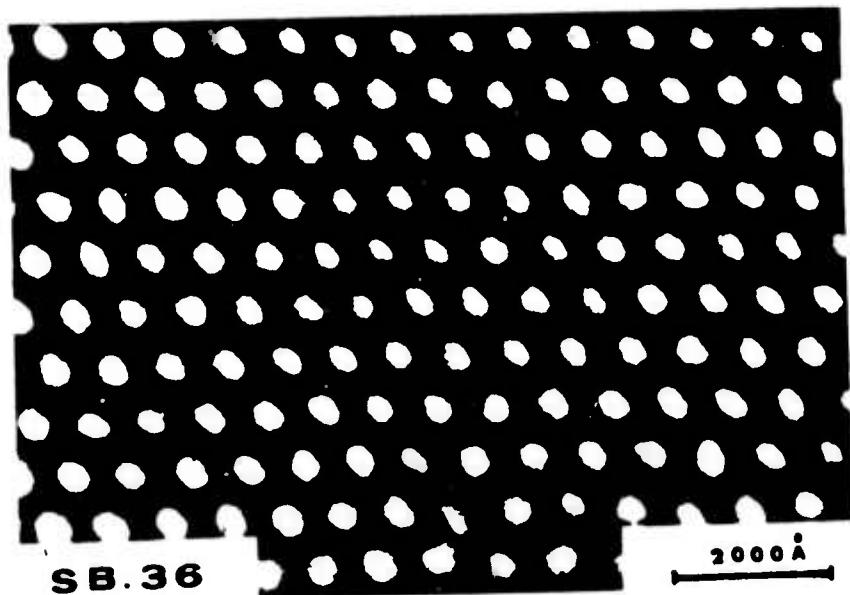


Figure 5. Electron micrograph of an S-B copolymer containing 69.5% S component showing "inverse" hexagonal structure.

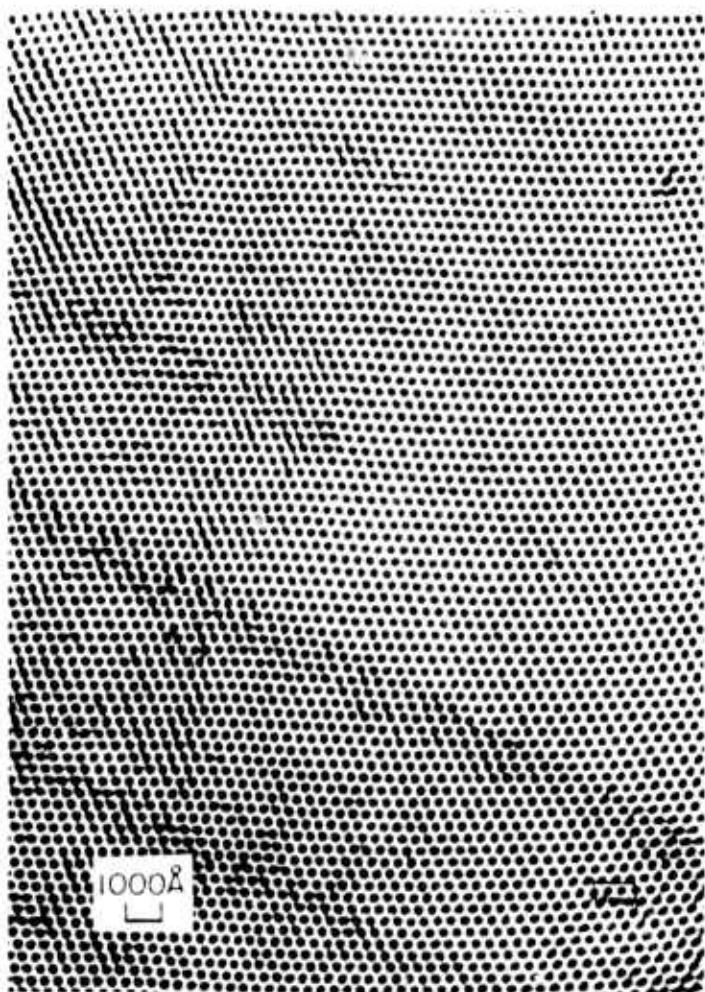


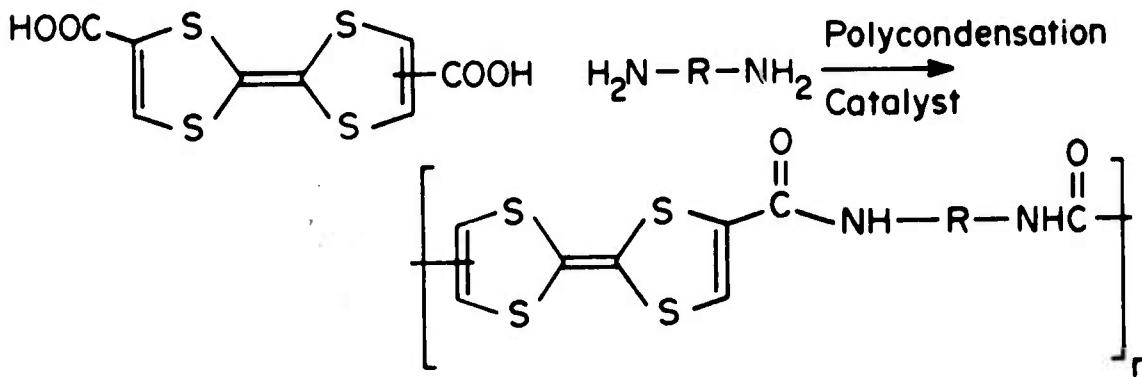
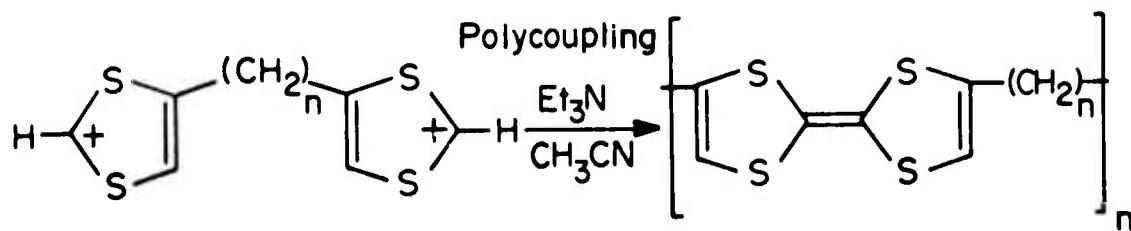
Figure 6. Electron micrograph of a B-S copolymer (68% S) showing long range order. Ultra-thin section.⁶

(i.e., organic metals) the situation is not so optimistic. First one must imagine the synthesis of an A-polymer which maintains a reasonable conductivity. This would require conduction along the macromolecule's chain (i.e., $(SN)_x$) or along a regular array of stacks (i.e., TTF-TCNQ). Building this regular structure into a block copolymer (or a homopolymer for that matter) appears difficult. Secondly, the interface between the A-domains would have to be such that, if many of them existed in the film, they would not be insulating. Probably, a semiconductor would result from this interface even if a given A-chain were a good conductor. For these reasons, the block copolymer concept appears to be a possible route to anisotropic semiconductors, but not to anisotropic conductors.

In Fig. 1 and 2 rather random arrangements of A and B domains are pictured. Actually, certain block copolymers are now known to have a very regular arrangement of the domains. Phase segregation occurs on a microscale, which for the usual systems studied is typically in the order of a few hundred angstroms. In microscopic mixtures, the scale can be much larger. The segregated microphases can be of three kinds⁴; spheres, cylinders, or lamellae as shown in Fig. 3.

Electron micrographs of styrene-butadiene block copolymers, for example, arrangement of these domains where styrene domains are surrounded by those of butadiene.^{5,6} Figures 4, 5 and 6 are representative. Thus, one can imagine a generalization of this phenomena to the copolymers proposed here.

Currently, no literature exists on block copolymers of semiconducting (or conducting) organic compounds. Further, the literature on organic metals incorporated into polymers is very scarce. Other than $(SN)_x$ only three reports exist. Bisfulvalenediiron - $(TCNQ_2)^-$ (bulk $\sigma = 10^2 \Omega^{-1} \text{cm}^{-1}$)⁷ has been incorporated into polymers via its 3-vinyl derivative.⁸ When 71% of the bisfulvalenediiron units were oxidized the conductivity was in the 10^{-2} to $10^{-3} \Omega^{-1} \text{cm}^{-1}$ range.⁸ Polymers of tetrathiafulvalene have also been synthesized via polycoupling⁹ and polycondensation¹⁰ procedures.



They have been complexed with TCNQ but no conductivity studies have yet been reported.

A key question which had not been answered is: What

portion of the conductivity of "organic metals" might be able to survive their incorporation into polymeric structures? Chemical and morphological order will be lost. Can better organic semiconductors result from the use of organic metals?

ACKNOWLEDGEMENT

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

References

1. L. Roth, presented at the ARPA Materials Council Meeting on One- and Two-Dimensional Conductors, July 10-11, 1975, La Jolla, California.
2. T. Garito, *ibid.*
3. R. N. Haward, The Physics of Glassy Polymers, John Wiley, New York, 1973. In particular see Chapter 8, J. Mann and G. R. Williamson and Chapter 10, M. J. Folkes and A. Keller.
4. C. Sadron, *Chim. Ind., Genie, Chim.* 96 (1), 507 (1966).
5. A. Douy and B. Gallot, *Molecular Crystals and Liquid Crystals* 14, 191 (1971).
6. G. Kampf, M. Hoffmann and H. Kromer, *Ber. Bunsenges physik. chem.* 74, 851 (1970).
7. U. T. Mueller-Westerhoff and P. Eilbracht, *J. Amer. Chem. Soc.* 94, 9272 (1972); D. O. Cowan and C. Levanda, *ibid.* 94, 9271 (1972).
8. C. U. Pittman, Jr. and B. Suryanrayanan, *ibid.* 96, 7916 (1974).
9. Y. Ueno, Y. Masuyama and M. Okawara, *Chem. Lett.* 603 (1975).
10. C. U. Pittman, Jr. and M. Narita, in press *Synthesis*, 0000(1976), reported at the Conference on the Electrical Properties of Polymers, New Paltz, New York, May 1975.

CHEMICAL ANCHORING OF MILDEWCIDES IN PAINTS:
AN ANALYSIS AND DISCUSSION OF ITS FEASIBILITY

C. U. Pittman, Jr.

ABSTRACT

Mildew defacement of organic coatings is a serious, wide ranging problem which has received exhaustive industrial attention. A wide variety of mildewcides have been used in paint systems. However, biocides have not been chemically attached to the paint's polymer system. By anchoring biocides to the polymer system of paint in a way that it may later be cleaved under biological (bacteria, fungi, etc.) attack, a paint that "fights back" might be prepared. A comprehensive literature survey was performed to assess the state of anchoring biocides in paints. Surprisingly, this has not been reported for paints used above ground. A critical review of the feasibility of this approach to the mildew defacement of organic coatings problem (MIDOC) was written. Experimental approaches for achieving this objective were outlined.

CHEMICAL ANCHORING OF MILDEWCIDES IN PAINTS.
AN ANALYSIS AND DISCUSSION OF ITS FEASIBILITY

C. U. Pittman, Jr.

PROCEDURE

After a brief initial review of the literature had been performed, a literature search methodology was developed. It included a consideration of the following topics, all of which must be considered when searching for evidence on whether or not biocides may be chemically anchored to paint and still be effective.

- A. Fungicides Chemically Anchored in a Coating
- B. Mechanisms of Fungicide Reaction with Mildew or Fungus
- C. Mechanisms of Mildew or Fungus Reaction with Paint or Painted Surface
- D. Reaction of Fungicide with Paint or Painted Surface
- E. Fungicides, Bactericides, Biocides, Mildewcides in Use with Paints, but not Chemically Bound
- F. Patents of Fungicides, Mildewcides, Biocides
- G. Reviews and Bibliographies
- H. Tests of Fungicidal Paints

Next, a description of each area was formulated along with a list of keywords. Then the Knowledge Availability Systems Center (KASC) at The University of Pittsburgh assisted in carrying out the search. It should be noted the KASC search concentrated only on recent literature, and hand search provided a coverage of earlier work.

RESULTS

The literature search has identified about 700 related references. Of these about 350 have been identified as having indirect potential value. Only a few have been sited here. Only a very few examples have been found where a biocide has been directly anchored to the coating material. This work has mainly been concerned with antifouling marine coatings. It has attracted wide attention recently, including a Chemical and Engineering News feature article and a special symposium in the January 1975 issue of the Journal of Paint Technology. To our knowledge, no reports exist where a house paint uses a vehicle to which a fungicide has been chemically attached.

The broad problem of Mildew Defacement of Organic coatings has been comprehensively reviewed by Brand and Kemp (Battelle Institute) for the Paint Research Institute.¹ A review of that breadth or magnitude was not attempted. Rather, a few narrow aspects of the MIDOC problem are covered in order to infuse views on coatings with anchored biocides into the midst of this broad MIDOC area.

The report is in two sections. Only the first is presented here. It consists of an analysis of the literature where some key research opportunities and needs have been emphasized. This section includes a discussion of anchored biocides and some pros and cons of this approach. Also we point out where a knowledge of the mechanisms of biocidal action would be useful to a consideration of anchored fungicide concept.

The second portion is a list of references (and in many cases the abstracts) of a large series of references uncovered by the KASC search entitled, Mildewcides in Paints, Project Profile 1-4550. One finds the references categorized according to the functional groupings listed as A-H on the first page. In addition it contains a preface which explains the purposes of the search and how the analysis was organized. A second, short analytical summary contains a pointed overview of the document surrogates relevant to the purposes of the search. A careful reading of this KASC document will uncover some useful references for almost any paint chemist. Furthermore, this document could be put to good use by those who may have no interest in the concept of biocide anchoring.

FUNGICIDES (BIOCIDES) CHEMICALLY ANCHORED TO COATINGS

The idea of attaching mildewcides, or biocides generally, to regular indoor and outdoor coatings has been virtually unstudied. Despite our extensive literature search, no published study was uncovered where one or more mildewcides was chemically attached to an outdoor paint's polymer system,*

*In a personal communication from Dr. Driscoll, Civil Engineering Laboratory, Oceanographic Structures Division, Point Hueneme, California, we have learned that research has just begun on the incorporation of tributyl and tripropyltins, via ester linkages, to alkyl based paints. These paints have been cured onto test boards. A. pullulan spores did not germinate on these samples in laboratory tests. Field testing is in progress. A report of this work will be forthcoming during the fall of 1975. Dr. Driscoll's work, to date, encourages the testing of concepts suggested in this review.

although a few inherently biocidal polymers were themselves studied as coatings. What key ideas have dominated thinking in this area which resulted in so little work on biocide anchoring? There are several such reasons.

1. In order for a chemically anchored biocide to be effective, it first must be demonstrated to be effective in a paint coating (when normally formulated by mixing). Then, and only then would it be anchored into the polymer vehicle.

2. If the biocide(s) are active when added to the paint, it is possible that they would be inactive, or less active, when chemically anchored to the polymer. For example, the use of anchoring could cause far less biocide to be available to the developing surface organisms under many circumstances (compared to the amount that would be available when the biocide(s) were not anchored). Thus, it could be argued that there is not any reason to pursue this approach because decreased activity might result.

3. It is thought, by many, that climax growth by A. pullulans may develop in some cases without actual damage to the paint vehicle. For example, growth could blossom from dirt caught in small crevases and cracks. The organisms could then collect their own food sources.² This view is somewhat controversial, but it must be considered. Thus, mildew growth could conceivably proceed without attack on the polymer containing the anchored biocide. Thus, the anchored biocide might be ineffective. However, it is also possible that such

an anchored biocide might act to inactivate exuded metabolites of A. pullulans or other necessary pre- or coinvasive species. If this inactivation of exuded metabolites actually interfered with life processes, growth suppression might still be achieved. Another line of thought that critics of biocide anchoring might pursue involves the possibility that defacing organisms could attack and metabolize cellulotic thickeners or other carbon sources in the film to which biocides were not anchored.

While the above possibilities exist, it should be remembered that many cases of damaging fungal growth, accompanied by bacterial attack and degradation, usually involve a continuous oxidative process by which the oil paint film is broken down.³ The fungi and/or bacteria actually attack the polymerized oil or binder weakening the film and forming acids. UV degradation probably is enhanced and release of free fatty acids further stimulates microbial attack.

4. Suppose that experimentation did prove that an active paint biocide was still active after anchoring. The paint chemist then might ask, so what? What is the point of carrying out the anchoring step as long as one could simply mix in the biocide in the first place? Wouldn't it be more expensive to anchor? Wouldn't it interfere economically with out current practice? These, indeed, are good points to consider.

5. If advantages were actually found and demonstrated for attaching biocidal groups to polymers for coatings, would

anyone bother to make such coatings anyway? Consider that many paint companies often do little polymer research. Biocides built into the structure of various monomers may cause large changes in the reactivity of these monomers. This would necessitate research efforts to find how to achieve the proper concentration of comonomer in the polymer vehicle. How could one tailor the molecular weight distribution, and how would one maintain the proper solubility and compatibility criteria? Thus, the general acceptability of such an approach has potential problems which would have to be faced by several layers of companies involved in paint production.

No literature now exists in which the topic of polymer-anchored fungicides is assessed directly. Thus, one is forced to consider the broad aspects of the total problem and make educated arguments from that vantage point.

SHOULD RESEARCH BE CONDUCTED ON POLYMER ANCHORING OF BIOCIDES?

In view of the above arguments, is there a good reason to do basic research in this area at all? Arguments can be made for answering with an unqualified yes. First, if a biocide, which is effective on simple mixing into the paint, is also effective when chemically attached to the polymer when dried into a film, we have uncovered an important clue. To be effective as a killer or growth inhibitor, that anchored biocide must have then been cleaved from the paint polymer. This cleavage must have been induced by microbial action (provided

it is shown that once the film has cured, water or solvents do not separate the biocide from the film). This fact alone could have important diagnostic implications.

Consider a situation where three different species of micro-organisms are found to be present, and even necessary, in succession during microbial attack and defacement. Call these species A, B and C, respectively. Now suppose that biocide X is effective in preventing climax vegetation of C when it has been mixed in a paint. One could then test X individually against A, B and C. Then the species which X retards (it may be more than one for a broad spectrum biocide), would then be tested on the paint with anchored X. If X is active when anchored, against let us say B, we would then know that the organism B has a way to cleave biocide X from the polymer. We would also begin to define one approach which would kill or retard C (climax organism) simply by interfering with the previous or coinvaders A and B. Alternatively, if X is biocidal to B when mixed, but not biocidal when chemically anchored, that gives us a lead as to how the biocide must be made available to the organism. It is important to accumulate this type of information on an expanded scale. It should be recognized that experiments using anchored versus non-anchored biocides provides the mildew research effort with one probe to study how the biocide must be available to the organism. This information could then be integrated with a knowledge of the microecology, with succession studies, and with surface enzy-

mology to allow us to develop a more detailed picture of how we might retard growth.

Other advantages of anchoring biocides are possible. Higher concentrations of killing agents might be permitted due to lack of leaching of the agent from the paint. Improvements may occur in the area of compatibility. Once anchored to the polymer vehicle, one could imagine producing a paint with a higher weight percent of biocide yet good mechanical properties. Also, the biocide would be dispersed in an even fashion and small chunks of pure biocide would not agglomerate or be found in the dried film.

Certain biocides are reasonably volatile. This is particularly true for nonionic examples. Even a small volatility can result in a serious loss from the paint film when one considers the high surface area, small thickness, and high exposure temperatures of the paint film in use. The volatility of the biocide may also pose a problem of toxicity. Some candidate compounds are so toxic that the paint film can be poisonous, or in service the paint film could release a toxic vapor. Chemical anchoring would remove or reduce this as a problem. Then the only toxicity problem would occur at the location the paint polymer was being produced. The dangers due to skin or lung exposure during application would be removed.

The addition of a mildewcide to a particular formulation can interfere with paint viscosity. Paints must be carefully formulated to avoid this problem. Thus, this can prevent

loading a biocide to the high levels desired in some cases. If the mildewcide is part of the polymer, then one could avoid this problem. The upper limit of polymer use for a particular formulation, as well as the weight percent mildewcide in the polymer, would then become the limiting factors in formulation.

EVIDENCE SUGGESTING DEFACING ORGANISMS CAN AND DO ATTACK PAINT POLYMERS

Ross⁴ demonstrated that bacterial and fungal deterioration of paint films adjacent to its substrate involve the following principal causative agents: A. pullulans, Cladosporium sp., Pseudomonas sp., and Flavobacterium sp. He stated, each of these organisms was capable of hydrolytic and oxidative deterioration of polymers containing ester linkages. Further, Ross⁴ observed that bacterial deterioration of the polymer binder at the paint-substrate interface contributes significantly to house paint failure and peeling. This general theme of a bacterial contribution to paint failure, as well as a bacteria-fungi symbiosis, has been echoed by many other workers.^{5,6}

The corrosive action of metabolic products of these organisms further degrades the film. It is possible these film decomposition products become food sources, or at the very least, a portion of the fluid media surrounding the organism.⁴⁻⁶ The ability of several species of fungi to hydrolyze fats, oils, fatty acids and linseed oil, presumably by exocellular enzymatic activity, has been noted by several bacteriologists.⁷⁻¹¹ Ferguson⁶ measured surface hydrolytic

and oxidative activity of fungi on paint and discussed, briefly, the enzymatic and metabolic pathways involved in the destructive process.

The food intake by fungi cells consists of the entry of dissolved substances into the fungi's protoplasm through cell walls. Macromolecules themselves won't pass these walls, but must be hydrolytically or fermentatively broken down. Further, the intake of toxic substances proceeds by the same path. The effect of toxic substances is dependent on their concentrations in the coating medium. Schmitt¹¹ suggested that the best fungicides must be simultaneously oil and water soluble.

From the above discussion, it would seem feasible that polymer anchored biocides in a paint film could be cleaved. In particular if the anchoring group were an ester linkage, hydrolysis seems likely. Alternatively other functional groups susceptible to hydrolysis or oxidative degradation should permit cleavage by micro-organisms. Since the concentration of active biocide is an important consideration, one might suspect that the combination of one or more resin-anchored biocides with a biocide "simply dissolved" in the paint might provide an advantage over the use of anchored biocides alone.

FUNGICIDES WHICH HAVE BEEN CHEMICALLY ANCHORED TO A COATING

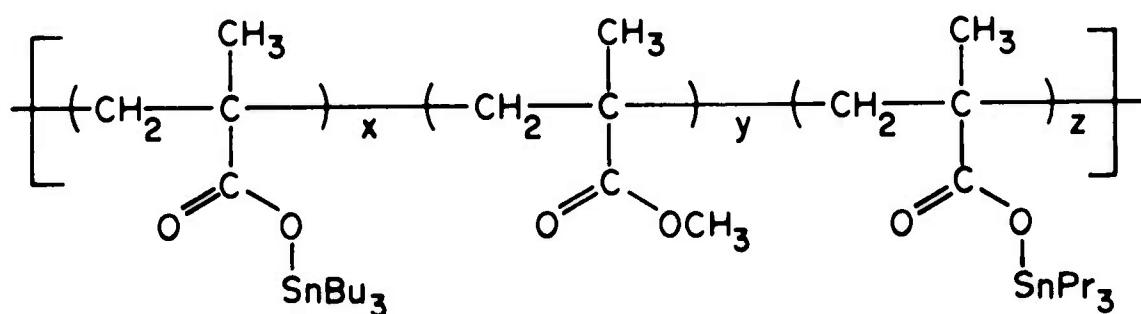
When one looks beyond antifouling marine coatings, the number of polymers prepared with anchored fungicides is small indeed. Ionene polymers (polyammonium salts) would be expected

to have strong antimicrobial and antifungal properties based on the action of quaternary ammonium fungicides. Indeed, this is the case and the antifungal properties of these polymers have been compared with their monomeric analogs. However, no data easily convertible into a meaningful picture for the paint chemist is yet available. A series of organometallic vinyl polymers has been prepared by Pittman et al.^{12,13} When cyclopentadienyltricarbonyl manganese was present in linseed oil or acrylic polymer films, antifungal properties were noted.^{13,14} However, these tests cannot be meaningful extrapolated to paint situations. Russian and Dutch workers have prepared a variety of biologically active organotin and organolead polymers and copolymers and other work on polymers for use in microbe-resistant clothing has been done. However, these uses are not related to paints. They do show the feasibility of synthesizing such polymers however.

A huge literature now exists on antifouling paints and coatings. For example, the January 1975 issue of J. Paint Technology was completely devoted to this topic and a search of the KASC listing, revealed many many references. Many types of antifouling coatings contain a biocidal agent tied into the polymer structure. Inorganic zinc coatings derived from the incorporation of zinc dust into silicate binders and self-curing inorganic zinc coatings have been made and reviewed.¹⁵ Chlorinated rubbers have been blended in a variety of anti-fouling paint systems.¹⁶ However, the antifouling coating

technology which is the closest, in concept, for transfer into regular paints is the development of polymer-anchored organotins.¹⁷⁻²²

A large series of polyacrylates, polyesters, epoxies and polyvinyls tested using monomers containing trialkyl and triaryltins, have been made.¹⁷⁻²² A typical polymer is that shown.



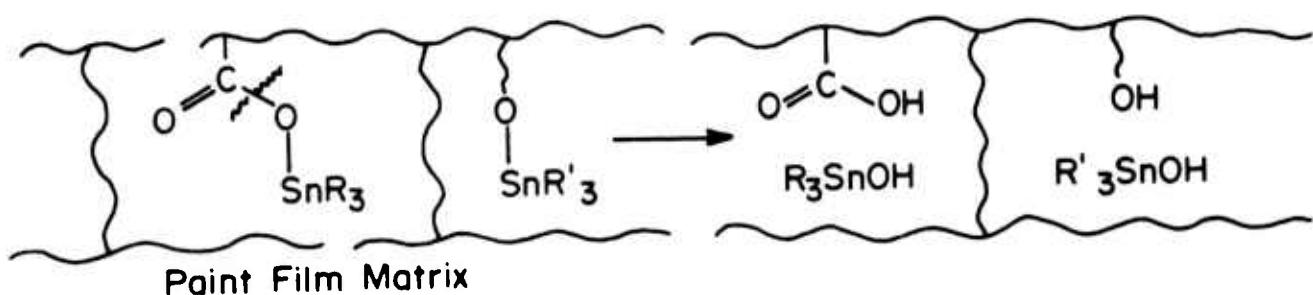
The biocidal agents which retard slime formation, and ultimately barnacle attachment, are R_3SnOH and their oxides. By incorporating these biocides into the polymer they are hydrolytically released and not simply leached away by mechanical action and solubility considerations. Thus, loss rates of the biocide were demonstrated to be 0.1 to 0.01 that of similar coatings where the organotins were simply added. Similarly, the loss rate was one order of magnitude less than that of currently used copper salt-based antifouling coatings.

These organometallic polymers have exhibited excellent antifouling performances after more than two years of exposure to severe tropical fouling conditions in Pearl Harbor. By conserving antifouling agents within the polymer matrix, the

organotin polymers offer the advantages of longer service life and negligible environmental impact.

What is particularly interesting to the paint chemist is that organotins have long been known as broad spectrum biocides for bacteria, algae and fungus.^{23, 24} The biocidal properties of organotin polyesters and polyethers have even been reported.²⁵ Incorporating two or more different R_3SnO -moieties into the polymers widens the kill spectrum of these polymers. Finally, organotin systems have been used as wood preservatives, particularly in Britain.

It is possible to begin to draw a parallel between this class of antifouling coatings and potential biocidal paint polymers for use on land. A paint polymer, which had anchored R_3SnO -groups, might "fight back" under conditions where ester and ether bonds were hydrolytically cleaved. This might be induced by an invading organism or might occur slowly due to environmental stresses.



The resulting R_3SnOH -derived products would remain in the film (if not injected by organisms) just as well as if it had been added to the paint originally. If it were to leach

or vaporize from the paint slowly, only that portion which had hydrolyzed could be lost. Thus, the tin level in the film would remain higher than would be the case if it was simply added to the paint as a monomeric additive during formulation.

The above discussion should not be considered as an argument for organotins per se, but instead it should serve to point out the potential for attaching several classes to biocides to paint polymers.

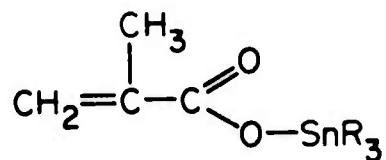
DESIGNING A MIDOC PROGRAM FOR ANCHORING BIOCIDES TO POLYMERS

Consumer paints which are frequently attacked by microorganisms are generally those used for painting interior or exterior house surfaces or commercial buildings. They include water-thinned paints and solvent-thinned paints. Water emulsion paints use film forming resins (frequently acrylics) or oils in water, while solvent-thinned paints employ drying oils or other oleoresinous materials. Thus, it would appear advisable to anchor biocides both to acrylic polymers and to oils.

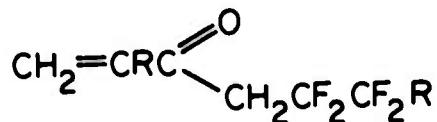
Neither concept is difficult. An appropriate vinyl (including acrylics) monomer of a biocide would be synthesized and then copolymerized to the desired stoichiometry into the prepolymer from which the latex is made. Also, appropriate vinyl derivatives will become chemically attached to linseed oil and other such oil films during drying. For example, Pittman previously demonstrated that vinylferrocene and vinyl-cymantrene when dissolved into linseed oil and dried to a film,

actually became part of the film.¹³ Solvent extraction of the pulverized film did not remove those chemically attached organometallics.

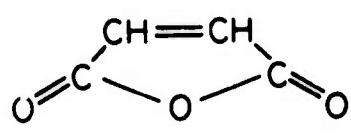
Some sample vinyl (and acrylic) monomers based on compounds with known strong antifungal properties are shown below. Most of these compounds are based on biocides which are EPA registered and have previously been used and recommended in paints of wide variety. They also have both fungicidal and bacterial activity.



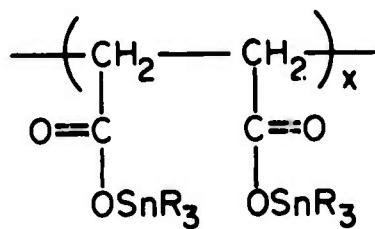
1 R = Bu, Pr



2

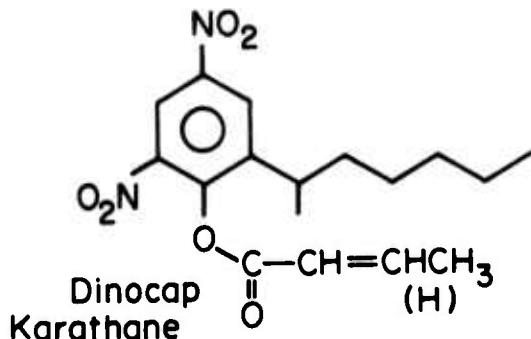


maleic anhydride

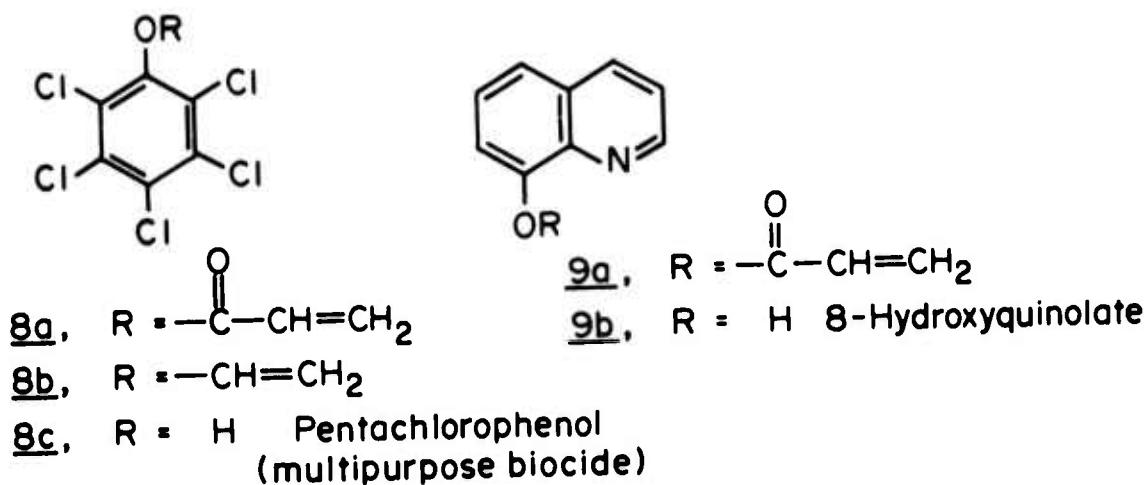
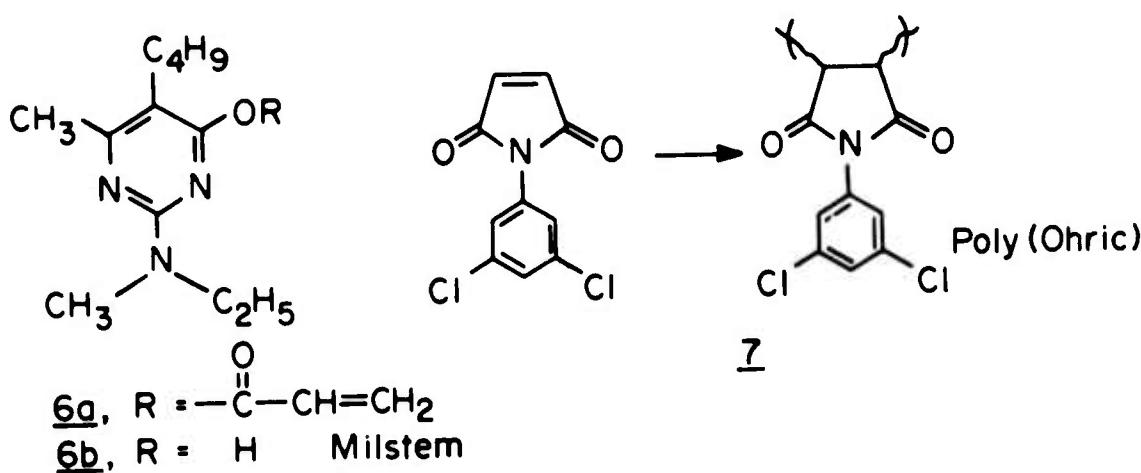
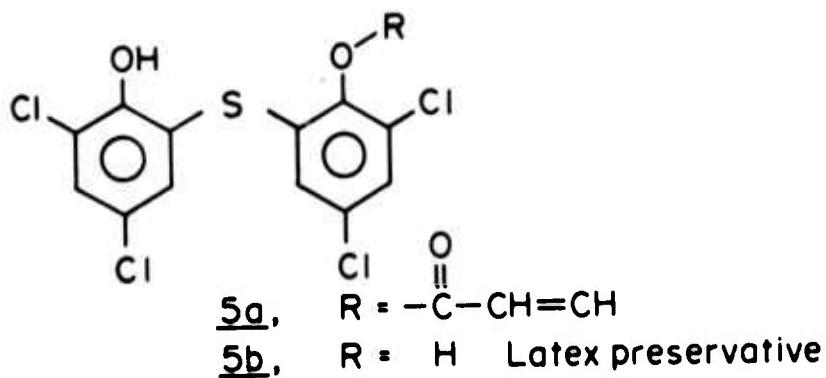


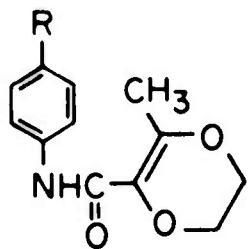
Grafted biocide

3



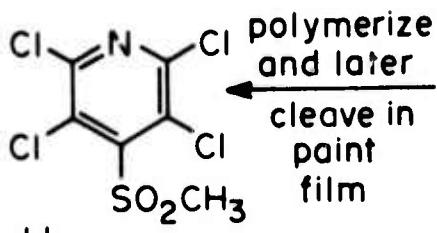
4



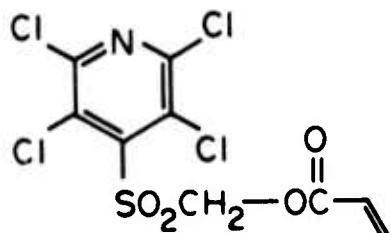


10a, R = $\text{CH}_2=\text{CH}-$

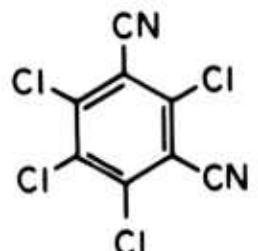
10b, R = H Vitavax



11 Commercial
Mildewstat
Dowicil 513



12 Polymerizable
Monomer

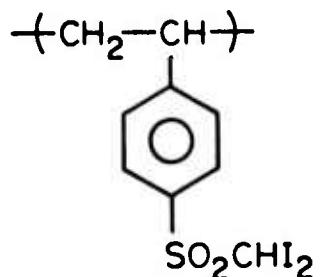


13 NOPCOCIDE

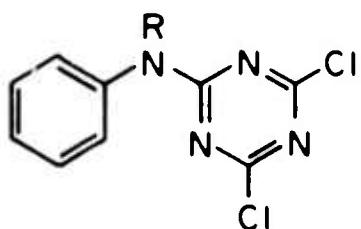


14a, Amical 48, R=CH₃

14b, R = -CH=CH₂

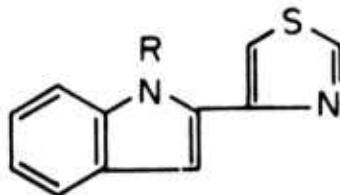


15



16a, R = $\text{CH}_2=\text{CH}-\text{C}-$

16b, R = H Dyrene

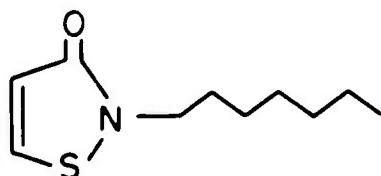


2-(4-thiazolyl) benzimidazole
Metasol TK - 100 where R=H

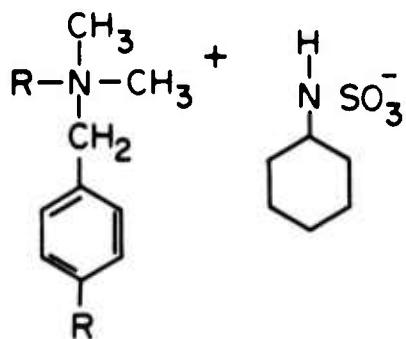
17a, R = H

17b, R = $\text{--C}=\text{O}-\text{CH}=\text{CH}_2$

effective only against fungi



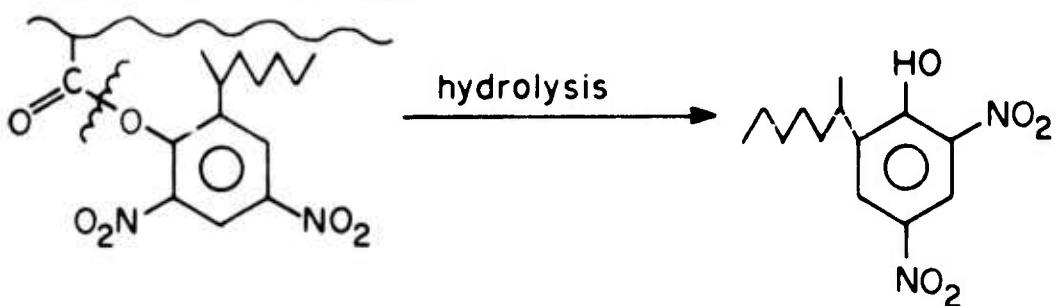
18 2-n-octyl-4-isothiazolin-3-one
Skane M-8



19a, R = Ethyl, Onykide 172 (ref. 27)

19b, R = vinyl

Dinocap 4 or its hydrogen analog could be copolymerized with traditional acrylic monomers to form a paint vehicle. This vehicle could be formulated into standard paint or the polymer could be cured as a film alone. In either case, standard laboratory and outdoor testing could be carried out on this polymer. In this case, we envision the following decomposition, induced by micro-organisms, could occur to free phenolic derivative 20.

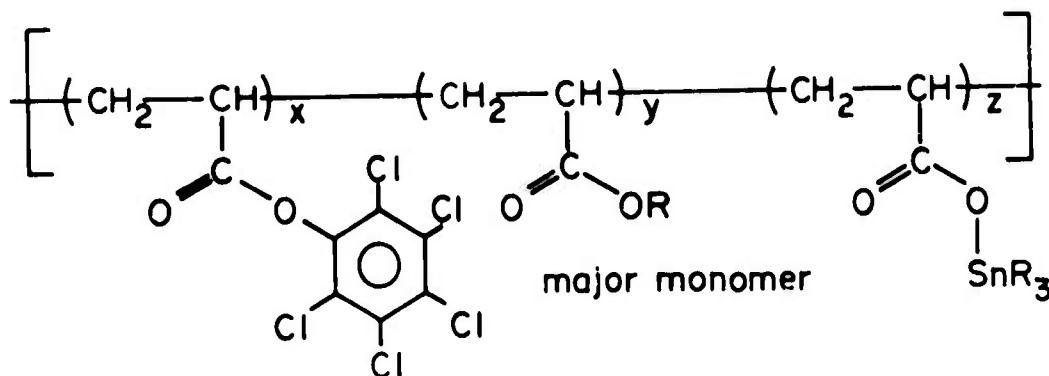


20 (incorporated
into organisms)

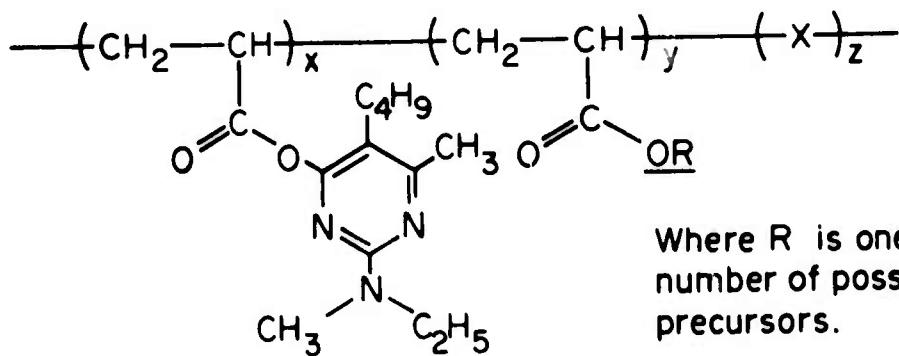
A paint or pure film of this type could be compared to a blank sample and another sample with an equal molar amount of the appropriate 2,4-dinitrophenol added. The short and long term protection afforded the coating by this fungicide in these forms would be compared.

It is recognized that compound 4, itself, might be a bad choice for protecting paints in actual practice. It is simply listed here to illustrate the type of route envisioned.

To make the most effective biocidal polymer, two or more biocidal monomers should be used to widen the kill spectrum. Thus, we envision an acrylic polymer such as:



biocidal monomer A
or



Where R is one of a large number of possible biocidal precursors.

A choice of which biocidal agents, or which appropriate combination of them, should be used would be made in consultation with those who have wide commercial and testing experience in the area. Furthermore, the addition of another biocide (such as a phenyl mercurial or any other suitable compound) to the paint by simple mixing would probably increase the effective mildew retarding ability of the system. Comparative tests of these possibilities should be carried out by standard methods.

CONSIDERATIONS REGARDING ANCHORING OF VARIOUS CLASSES OF
BIOCIDES ORGANOMETALLIC COMPLEXES

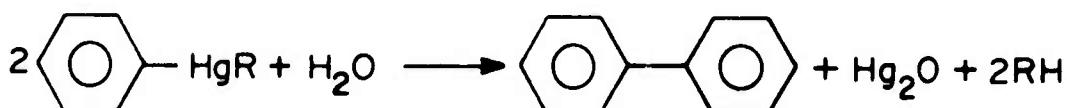
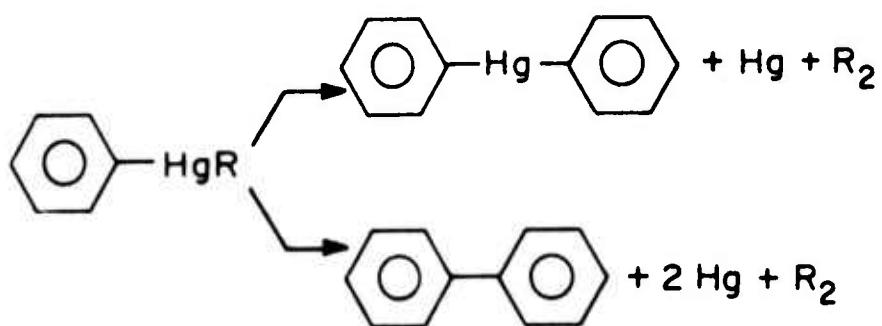
Organolead compounds are well known as broad spectrum paint biocides but environmental considerations currently rule against much research with these compounds, regardless of whether they are polymer-anchored or not. Organotin derivatives, on the other hand, are currently widely used in antifouling coatings for ships and bis(tributyltin) oxide, for example, is well established as a commercial timber preservative.²⁸ In the British Isles, organotins have become the preferred preservatives for exterior millwork and their formulations have been comprehensively reviewed.²⁹ These compounds are broad spectrum biocides active against fungi, insects, nematodes and bacteria. Thus, organotins have many of the advantages that organomercurials exhibit. Following an early report of Arnold and Clarke,³⁰ considerable interest in the application of tributyltin oxide as an antifungal agent in emulsion paints developed. The application of both organotin and organolead compounds in paints has more recently been reviewed by Giesen³¹ in an article of importance to read prior to attempting anchoring studies.

Organomercurials, in particular phenyl mercury salts, have long been standard, broad spectrum biocide used to prevent electron probe paint from defacing organisms. These compounds also give "in can" protection against microbial deterioration. The mechanism by which phenyl mercurials protect painted surfaces is highly controversial and has been the

subject of an enormous number of studies (for a review see Broome and Lowrey³²). For this reason, it is of interest to cover just a few aspects concerning the use of this class of compounds in order to compare these concepts to that of an anchored biocide.

First, phenyl mercurials migrate in films. Spectro-photometric analysis and electron probe microanalysis show mercury present in the wood under latex paints six years later.³² The anion has little influence on activity or migration rate.³³ Secondly, it has generally been noted mercurials give good protection for periods varying from 6 months up to 2 years but after this they are not very effective.

Hoffmann and coworkers³⁴⁻³⁶ have painstakingly studied the decomposition reactions of phenylmercurials in paint films, and they have related the mildew resistance of the paint to the decomposition of the phenyl mercurials. Gloss enamel paints showed a steady loss of mercury content (to 20% of its original value after 17 months) and the mildew resistance declines markedly during this period. Using known vapor pressures for the chloride and acetate at 30° and 60° and estimating the effects of film entrapment, it was determined that the observed losses were not due to volatility. The best current evidence is that both thermal decomposition and reaction with water (perhaps photochemically induced) occur as follows:



The metallic mercury may then migrate to the wood or vaporize. Data published by O'Neill³⁷ do not give much hope of finding phenylmercurials which have a long film life and this agrees with the results of Whiteley^{38, 39} and Dimond and Stoddard,⁴⁰ as well as those of many others.¹

These results emphasize one problem which polymer-anchoring cannot solve. If the biocide is inherently chemically labile, its attachment to a polymer will not prevent its chemical degradation. Thus, anchoring phenyl mercurials to polymers in order to maintain a high mercury content in the paint film is a concept doomed to failure. The same cleavage of the phenyl-mercury bond, illustrated above, would occur whether or not the phenyl mercury derivative was polymer anchored.

Research results of many groups suggests mercury compounds exert their effects on organisms by combining with enzymes,⁴¹ particularly with essential sulfhydryl groups in

enzymes.⁴²⁻⁴⁴ We have not encountered literature which specifically deals with the biochemical mechanism by which organotin compounds act on A. pullulans or other paint defacing organisms.

PH. An alkaline environment discourages mildew growth on paints.⁴⁵ This has been attributed to retardation of the metabolism rate of these organisms.⁴⁶ However, on the molecular biochemical level the way this takes place has not been defined, and it is an area of research which needs to be explored. Paints containing zinc oxide also exhibit a degree of mildew resistance which has been related to alkalinity and to its surface hardness as well as to other phenomena such as chaulking.

Miscellaneous. Paints containing a fungicide plus zinc oxide have been shown to resist mildew growth better than analogous paints containing only zinc oxide or the fungicide.⁴⁷ This was true when the fungicide was phenyl mercuric succinate, 2,3,4,6-tetrachloro-4-methylsulfonyl pyridine, or ethylbenzyl dimethyl alkyl ammonium cyclohexyl sulfamate. Such cooperative or symbiotic effects should also be manifest when biocides are anchored.

Often micro-organisms are already present on the wood substrate before the paint is applied or they can come in contact with the film during drying.⁴⁸ The importance of this fact in relating anchored versus "added" biocides is unknown.

FUNGAL MORPHOLOGY, METABOLISM AND REPRODUCTION

Fungi comprise a large, heterogeneous and ubiquitous group of heterotrophic organisms which live as saprophytes or

parasites, or associated with other organisms as symbionts.^{1,62} Fungi are characterized by a distinctive, filamentous, multi-nucleated vegetative structure, the mycelium, which in turn consist of a branching system of walled tubes, the hyphae. The mycelium may consist of a single multinucleate cell (acoenocyte) in which there are no transverse septa or it may be multicellular and composed of uni-, bi- or multinucleated cells. Nutrients are absorbed over the whole surface of the mycelium, thus forming a highly effective system for exploiting the available resources of the substratum. Growth proceeds for a period and is followed by reproduction giving rise to spores, uni or multicellular bodies that in turn give rise to new individuals.

A fungal cell may be considered to consist of two main parts: the outer firm cell wall, of varying thickness and constitution, and the inner protoplasm. The protoplasm is enveloped in a cell membrane and contains the various cell organelles. In some cells the cytoplasm varies in consistency and the viscosity may change under certain conditions.

Reproduction in fungi results in the formation of enormous numbers of spores, uni- or multicellular, microscopic propagules containing one or more nuclei, which are liberated from the thallus passively or by active discharge.⁴⁹ Typically both asexual and sexual spores are formed, and the corresponding stages of the life cycle are described as imperfect and perfect, respectively. The participation of hormones in the sexual

reproduction of fungi is a well-established fact.⁵⁰ Spore dispersal can occur in an enormous variety of ways. The spore content of the air changes with changing weather and at different times of day.^{51, 52} Davis⁵³ has shown that wetable spores of A. pullulans, (the major paint defacing organism), also are carried with raindrops.

Fungi can obtain their nitrogen source as an organic or an inorganic form. The fungi produce, in addition to CO₂ and water, all the complex materials required for cell structure and vital processes. This represents a large number of substances of varying complexity, whose function is still obscure.⁵⁴ Some of these metabolites, particularly exocellular products, may be involved in the mildew-defacement problem and should be examined.

Metabolism in fungi is extremely diversified.^{55, 56} There are, however, certain unifying pathways. The identification of fungal products was a necessary preliminary to understanding their mechanism of biocytogenesis. Since fungal products are large and highly diversified, it is often convenient to classify such products on the basis of chemical structure, even though difficulties arise, because many products fall into more than one category. A classification should be made of metabolites of A. pullulans.

In addition to the special chemical products involved in biosynthesis and in the production of "secondary" metabolites which are often secreted into the substratum, the metabolism

of nutrients consumes a large part of the intercellular biochemistry of fungi. However, with regard to the species of interest, A. pullulans, little information exists in the current literature. One must rely on the early, but limited, reports of Clark and Wallace⁵⁷ for specific information regarding metabolism in A. pullulans. Again this work is limited to carbohydrate metabolism. They showed that a strain of A. pullulans possessed constitutive enzymes for the utilization of several sugars; they also reported that young cells of the organism were capable of oxidizing glucose through the pentose phosphate cycle and that older cells use, in addition, some of the reactions of the Embden-Meyerhof system.

Studies with intact cells and cell-free extracts of A. pullulans showed the presence of the necessary enzymes for oxidizing most of the compounds of the Krebs cycle. Adaptive enzymes were present under certain conditions for the oxidation of acetate. When cells were buffered at pH 6.8, acetate and ethyl alcohol were oxidized at rates of more than 1/2 and 1/3, respectively, that of glucose. α -Ketoglutarate, fumarate, cis-aconitate, and succinate were slowly oxidized; the citrate, isocitrate and malate, however, were not oxidized.

One must turn to the work of Winters⁵⁸ for a more recent biochemical examination of A. pullulans. Winters has shown the production of certain extracellular enzymes including cellulose by this organism. However, an earlier report⁵⁹ states that even the frequent occurrence in nature on a cellu-

losic substrate, is distinctly not acceptable evidence of ability to utilize directly cellulose. A. pullulans is an organism frequently isolated from exposed cloth which is not cellulosic.⁵⁹

THE NEED FOR A KNOWLEDGE OF HOW BIOCIDES ACT BIOCHEMICALLY

In a review of the mode of action and relation of chemical structure to activity of paint fungicides, Torgeson⁶⁰ stated, "Unfortunately, knowledge of fungal biochemical processes is so meager that it is not possible to design superior fungicides based on rational principles." Another point to be made is that many of the fungicides used in paints today were first used as agricultural compounds in plant protection. Many have relatively low toxicity and as such often exhibit a fairly narrow margin of selectivity when employed to kill or inhibit germination of fungi. Since selectivity is not a factor in protecting a paint, it would appear possible to develop more effective fungicides (i.e., fungitoxic compounds without the phytotoxicity limits demanded in agriculture).

Ritchie⁶¹ claimed that an important cause of failure of fungicides when used in coatings was the lack of a sufficient vapor pressure of the fungicide which would make direct contact with the spores. However, the need for a vapor pressure has not been conclusively demonstrated as the means of inducing spore to biocide contact. Other modes of contact are possible. Furthermore, for each different biocide, the

biochemical and metabolic action which leads to its suppressive effect could be different. Clearly, this is an enormous field.

Investigators who want a comprehensive survey of fungicides, to include their agricultural and industrial applications, environmental interactions, and chemistry and physiology, are referred to the advanced treatise edited by Torgeson.⁶² Extensive review chapters covering fungitoxicity, commercial fungicide development, formulation, application, industrial applications, wood preservation, chemical interactions, microbiological detoxication, mechanisms for reaching the site of action, as well as each of the major chemical classes of fungicides have been written. Despite this enormous effort, the equivalent of only about two pages (out of a total of 1440) exist concerning the action of fungicides in paint films! If this void could be even partially filled, the design of mildew resistant paints would be aided.

Currently, the empirical generalization exists that increased protection is afforded a paint by the use of multiple biocidal agents. Suppression of such a wide range of pre- and coinvasive bacteria and fungi as well as the climax species would best be achieved by using several biocides in combination. The same would be true of polymer anchored biocides. Thus, paint polymers should be prepared with more than one bound biocide. For example, chlorinated phenols, organotins, and thiazolyl benzimidazole moieties could all be included in the same vehicle. Alternately, two biocides could be anchored and others

could simply be added to the paint. A variety of combinations may be envisioned. It is possible that suitable compatibility between hypothetical biocides A, B and C could be obtained (when normally this was not the case) if one or more of them were anchored to the polymer.

Finally, the work on organism succession being carried out by Schmitt⁶³ and Winters⁶⁴ is important to the design of biocidal films for several reasons. First, biocides should be found which are particularly active against the necessary organisms leading to climax vegetation. To do this we must know what these organisms are and what successions are involved in each climate zone and for each major class of paints. Then, the metabolic basis (mechanism) of biocidal action for several effective biocides should be worked out for each succession organism. When this is known, one could then begin the process of rationally designing a mildew resistant paint.

ACKNOWLEDGEMENT

This research was supported, in part, by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan.

References

1. B. G. Brand and H. T. Kamp, "Mildew Defacement of Organic Coatings. A Review of Literature Describing the Relationship Between Aureobasidium Pullulans and Paint Films, August 1973.
2. S. Shapiro, Official Digest, 622 (1962).
3. R. T. Ross, Official Digest 30, (399) 377 (1958).
4. R. T. Ross, "Microbiological Deterioration of Paint Films," Developments in Industrial Microbiology 6(14), 149 (1964).
5. J. Schmitt and D. Padgett, J. Paint Technology 45, 581, 32 (1973).
6. J. F. Ferguson, Paint Technology 33, (6) 18 (1969).
7. C. Duncan, Official Digest, October 1963.
8. D. Perlman and S. Wagman, J. Bacteriol. 63, 253 (1952).
9. F. D. DiAccadia, Chem. Abst. 51 (2), 1379 (1956).
10. K. Weise, "Attack on Linseed Oil Paints by Mould Growth," Paint and Varnish Production Manager 10 (6), 1934.
11. K. Meir and H. Schmidt, Paint and Varnish Production 46, 13 (1954).
12. C. U. Pittman, Jr., J. Paint Technology 43 (84) 29 (1971).
13. C. U. Pittman, Jr. and R. Hanes, ibid 45 (582), 78 (1973).
14. C. U. Pittman, Jr., G. Marlin and T. D. Rounsefell, Macromolecules 6, 1 (1973).
15. D. H. Gelfer, J. Paint Technology 47 (600), 43 (1975).
16. P. A. Herbert, D. F. Bowerman and K. S. Ford, ibid 47 (600) 48 (1975).
17. J. A. Montemarano and E. J. Dyckman, ibid 47 (600) 59 (1975).
18. A. W. Sheldon, ibid 47 (600) 54 (1975).
19. E. J. Dyckman, J. A. Montemarano and E. C. Fisher, Naval Eng. J. 33, December 1973).

20. J. A. Montemarano and E. J. Dyckman, Naval Ship Research and Development Center Report 4159, August 1973.
21. E. J. Dyckman, J. A. Montemarano and E. C. Fisher, Naval Eng. J. 59, April 1974.
22. E. J. Dyckman and J. A. Montemarano, Am. Paint J.. Aug. 20, 1972.
23. J. G. A. Luijten and G. J. M. vander Kerk, "Investigations in the Field of Organotin Chemistry," Middlesex England: Tin Research Institute (Oct. 1955).
24. A. Sijpesteijn, Antonie van Leeuwenhoek 28, 346 (1962).
25. S. Midgal, Isreal J. Chem. 5 (4), 163 (1967).
26. W. E. Higbee, Modern Chemicals, 1-14, March/April 1966.
27. S. J. Lederer, American Paint J. 55, 31 (March 1971).
28. A. G. Beaumont and C. A. Mackay, T.R.I. Publication No. 489, Internal Pest Control, May/June 1974.
29. B. A. Richardson, "Tributyltin Wood Preservatives," Rec. Ann. Conv. B.W.P.A. (1970). Reprinted Paperi ja Puu, 19-26, 1 (1971).
30. M. H. M. Arnold and H. J. Clarke, J. Oil Colour Chemists' Assoc. 39, 900 (1956).
31. M. Giesen, Metallorganische Verbindungen als Biocide fur Anstrichstoffe, F.A.T.I.P.E.C., Congr. 8, 185. Chemie, Weinheim Bergstr., Germany; also see G. J. M. vander Kerk and J. G. A. Luijten, J. Appl. Chem. 4 314 (1954), ibid. 6, 56 (1956).
32. T. T. Broome and E. J. Lowrey, J. Paint Technology 42, 543, 227 (1970).
33. G. Gassner, Phytopathol. Z. 17, 1 (1951).
34. E. Hoffmann and O. Georgouassis, J. Oil and Colour Chemists Assoc. 43(11), 449 (1960).
35. E. Hoffmann and B. Bursztyn, ibid., 46(6), 460 (1963).
36. E. Hoffmann and B. Bursztyn, ibid., 47(11), 871 (1964).
37. O'Neill, JOCCA 46, 425 (1963).

38. P. Whiteley, JOCCA 43, 842 (1960).
39. P. Whiteley, ibid. 45, 803 (1962).
40. A. E. Dimond and E. M. Stoddard, The Conn. Agriculture Station, New Haven, 1955, Bulletin 595.
41. H. Frank, Zentr. Bakterid. Parasitenk., Abt. 108, 666 (1955).
42. P. Fildes, Brit. J. Exper. Path. 25 76 (1940).
43. D. Gottlieb, Phytopathology 47, 59 (1957).
44. W. L. Hughes, Ann. N.Y. Acad. Sci. 65 454 (1957).
45. R. C. Trueblood, J. Paint Technology 40, 582 (1968).
46. P. H. Krumperman, "Microbiology of Paint Films, V.," Am. Paint J. 42, 72 (1958).
47. R. C. Trueblood, J. Paint Technology 43(554), 76 (1971).
48. J. F. Ferguson, Paint Technology 33(6) 18 (1969).
49. C. J. Alexopoulos, Introductory Mycology, John Wiley & Sons, New York (1964).
50. E. O. Wilson and W. H. Bossert, Rec. Prog. Hormone Res. 19, 673 (1963).
51. M. Mallea, I. G. Murray, G. Segretain, C. M. Philpot, H. Charrin, E. Gueho, and J. Charpin, Acta Allergologica 27, 273 (1972).
52. E. M. Dupont, R. C. Field, G. E. Leathers and W. T. Northey, J. Allergy 39(4), 238 (1967).
53. R. R. Davis, Nature 191, 616 (1961).
54. W. B. Turner, Fungal Metabolites, Academic Press, New York (1971).
55. V. W. Cochrane, Physiology of Fungi, John Wiley & Sons, New York (1967).
56. S. Shiba, S. Natori and S. Udagawa, List of Fungal Products, Charles C. Thomas, Springfield, Ill. (1964).
57. D. S. Clark and R. H. Wallace, Cand. J. Microbio. 4, 125 (1958).

58. H. Winter, Report of the Research Director Proceedings, Paint Research Institute 47(604) 33 (1975).
59. R. G. H. Siu, Microbial Decomposition of Cellulose, Theingold Pub., New York, p. 531 (1951).
60. D. C. Torgeson, J. Paint Technology 38(498), 368 (1966).
61. W. H. Ritchie, J. Paint Technology 16(191), 481 (1951).
62. D. C. Torgeson, editor, Fungicides, An Advanced Treatise Vol. I and III, Academic Press, New York, 1967 and 1969, respectively.
63. J. Schmitt, J. Paint Technology 45(581), 32 (1973); ibid. 46(599), 59 (1974).
64. H. Winters, J. Paint Technology 44(575), 39 (1972); ibid. 46(594), 69 (1974); ibid. 46(598), 49 (1974).

COMMENTS ON SYNTHETIC APPROACHES TO
ONE-DIMENSIONAL CONDUCTORS

C. U. Pittman, Jr.

ABSTRACT

The drop in conductivity exhibited by tetrathiafulvalene-tetracyanoquinodimethane (TTF-TCNQ) complexes below T_c appears, at least in part, to be a result of Peierls instability. This might be avoided by increasing interstack interactions or by locking the TTF-TCNQ stacks into a "super lattice" similar to graphite. Specific synthetic ideas along these lines are mentioned.

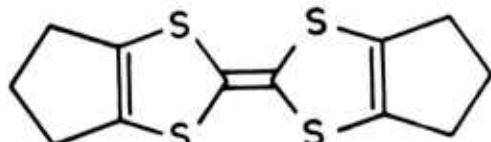
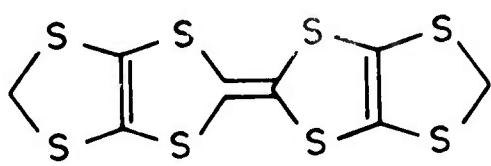
COMMENTS ON SYNTHETIC APPROACHES TO
ONE-DIMENSIONAL CONDUCTORS

C. U. Pittman, Jr.

Two key phenomena appear to limit the development of improved organic metals currently. These are the Peierls instability distortions and the magnitude of interchain interactions. For example, in TTF-TCNQ as the temperature is lowered below its T_c , 66°K, the conductivity rapidly decreases. This is thought to be due in part to the Peierls distortion. If strong interchain interactions were present, it is possible TTF-TCNQ would continue to be conducting to ~0°K or possibly show a transition to superconductivity. The interchain interaction would allow a path around the Peierls distortion. This could be a reason why $(SN)_x$ polymers remain conducting as they approach 0° (and become superconducting at 0.37°K). In $(SN)_x$, many feel that strong interchain interactions exist between sulfurs in adjacent chains which permit such behavior. Alternatively, Peierls distortion might not occur in this system. Thus, one synthetic approach which should be encouraged involves synthesizing systems with improved interchain interactions. Secondly, synthetic studies could be undertaken to lock chains into position in a manner to remove Peierls distortion. The latter could be difficult due to the small distances involved

in such distortions.

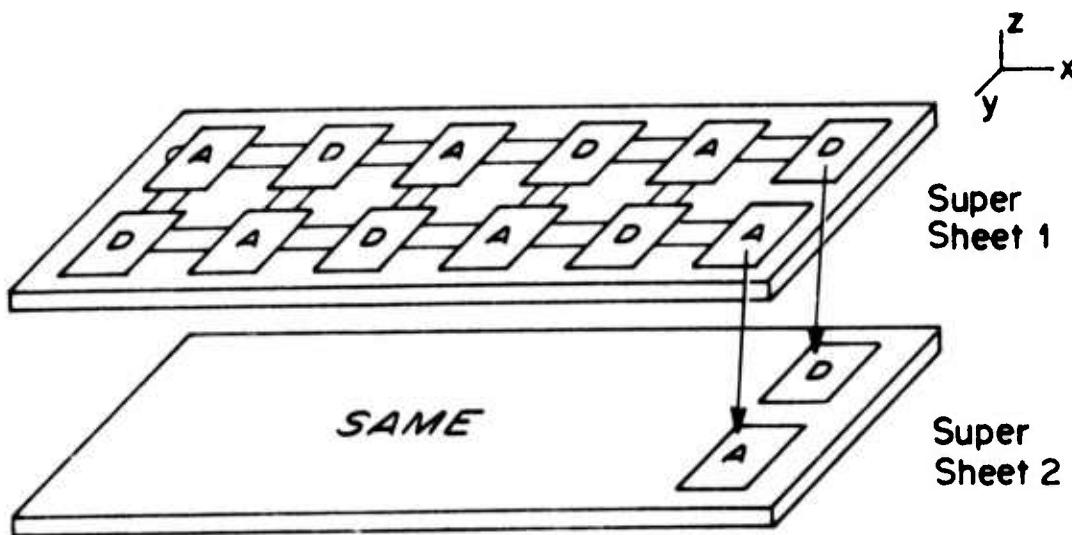
Several synthetic targets can be suggested such as TTF derivative 1. Since it is now known that TTF derivative 2 continues to remain a good conductor down to 0.037°K ,¹ serious



Peierls distortion did not occur. In 1 we have extra sulfur atoms, conjugated to the TTF system, which might force better interchain interactions with the nitrile nitrogens of the TCNQ stacks. The great difficulty which currently exists is our inability to predict closely the detailed crystal packing of such TCNQ complexes as 1 TCNQ. Only when the crystal structure of some very close analogs are already known can the structures of new materials be suggested. Thus, synthetic approaches should not be overly directed at present.

Increasing the mass of a group or incorporating it in a rigid lattice structure could prevent Peierls instability. Graphite sheets, for example, would be an example of this behavior. Perhaps one dimensional materials with greater rigidity could be synthesized in analogy to graphite.

¹A. Block, presented at the ARPA Materials Research Council Meeting on One- and Two-Dimensional Conductors, July 10-11, 1975, La Jolla, California.



Arrows indicate the conducting
z axis (σ_{max} axis).

As shown above, the super sheet is completely hooked together in two dimensions. Alternatively, hooking may only be along the X or Y axis. Synthetic schemes to produce such materials will require new imaginative chemistry, and it does not appear easy.

ACKNOWLEDGEMENT

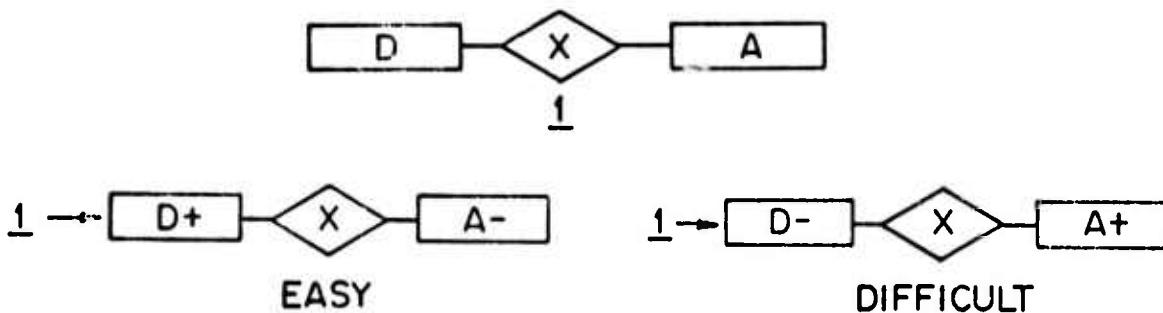
This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-0253 with The University of Michigan

MOLECULAR RECTIFIERS?

C. U. Pittman, Jr.

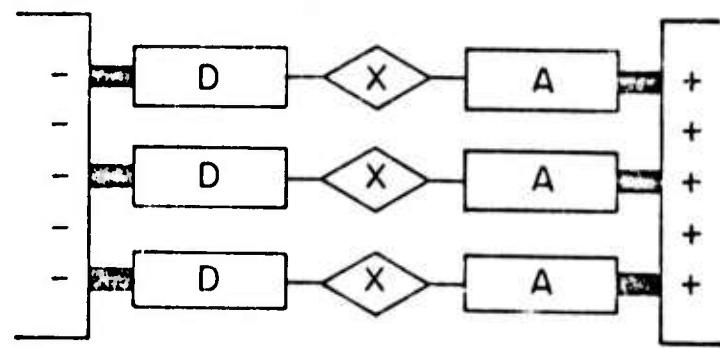
ABSTRACT

The synthesis of donor acceptor molecules (such as 1) with an insulating connecting group between them might be useful as rectifiers. Thus, current flow from donor to acceptor is much easier than the reverse case. By orienting such molecules



between "plates" or contacts, one can envision building a "molecular" rectifier such as that shown below which contains three donor-acceptor molecules at the interface.

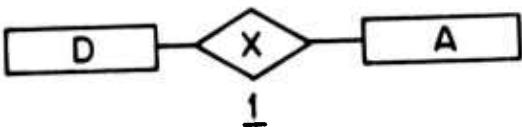
Calculations were performed on an example molecule of the TTF-TCNQ type, using net polarizabilities estimated for the molecule, to obtain I_R/I_{Rc} as a function of ρ_x , ω , and M . The frequency is ω , the number of molecules of type 1 was M , and ρ_x is the resistivity of the array in the direction of the donor to acceptor axis. It was shown that as long a M is kept small, such rectifiers could be useful for fairly low frequency applications.



MOLECULAR RECTIFIERS?

C. U. Pittman, Jr.

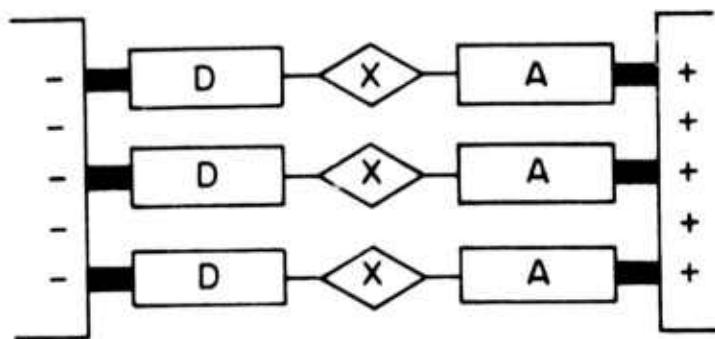
Synthesis of donor-acceptor units which are attached together in a rigid fashion could be used to generate a variety of other useful properties besides one-dimensional conductors. For example, in a hypothetical molecule, such as 1,



the donor portion is insulated from the acceptor portion by moiety X. Such a molecule could be used as a rectifier because it would be far easier to drive electrons from D to A across gap X than the reverse. Thus, if

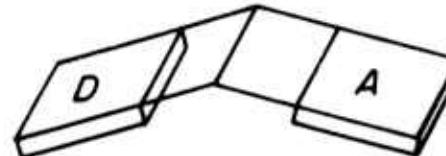
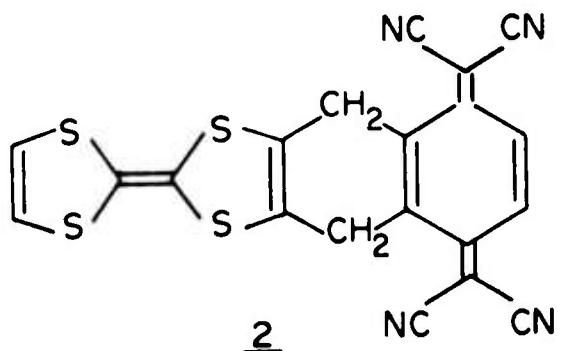


a series of molecules, such as 1, were oriented between plates as shown below, one would have the potential to build "molecular rectifiers" if the plates to which they were attached were of molecular size. Thus, only the technology of size reduction of the end plates would be limiting. Also a great size reduction is possible in thickness (i.e., distance between plates).



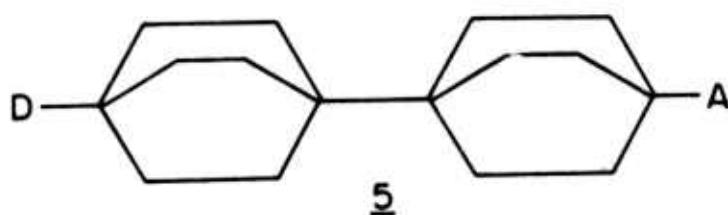
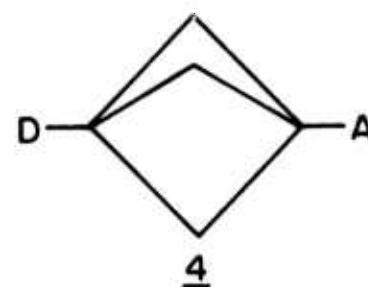
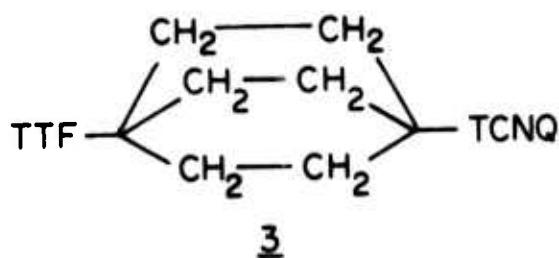
What molecules might be suitable for such an application?

Here the synthetic chemist can envision innumerable possible targets. Simple conceptually is 2. However, this molecule



BENT

could bend across the methylene bridges. To obtain rigidity, molecules such as 3-5 can be envisioned where the distance between D and A can be continuously varied.



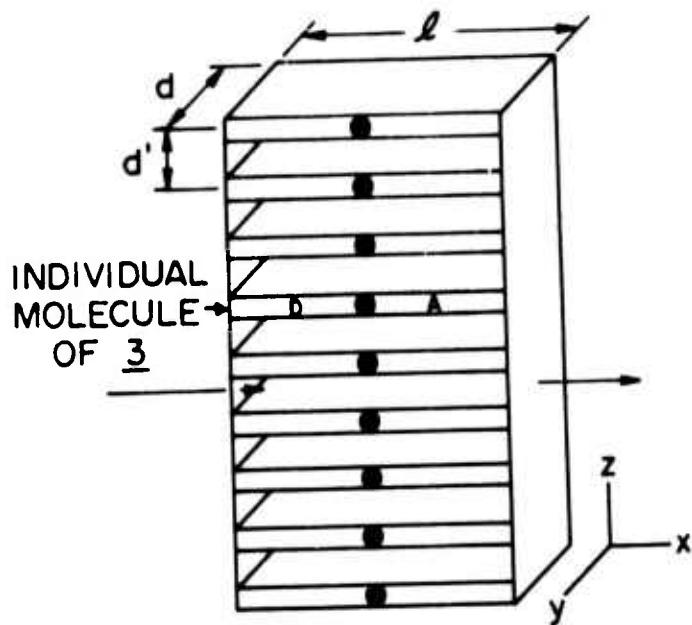


Figure 1.

C = Capacitance

M = Number of molecules in the linear stack

α = Polarizability per molecule cm^3

ρ = Bulk resistivity in esu across a given molecule. That is,

the resistivity perpendicular to the molecular stack and
between the plates (i.e., along the x -axis).

ω = Frequency of current in CPS.

Thus, the gap size (electronic) can be tailored by synthetic design. It may also be necessary to suitably derivatize the donor and acceptor to allow them to attach to the end plates or adhere there in a regular way. In any case, the concept is an interesting one which might bring high return.

In order to see if this concept was feasible for application over a reasonable range of frequencies, a series of sample calculations were made (with the aid of H. Reiss). The rectifier model is shown in Fig. 1. In this case the distance between plates is given by ℓ , which is also taken as the overall length of the "rectifier" molecule \underline{l} .

In the sample calculations, the simplifying assumption was made to neglect end effects. Since this rectifier is being considered on a molecular scale the following equations apply.

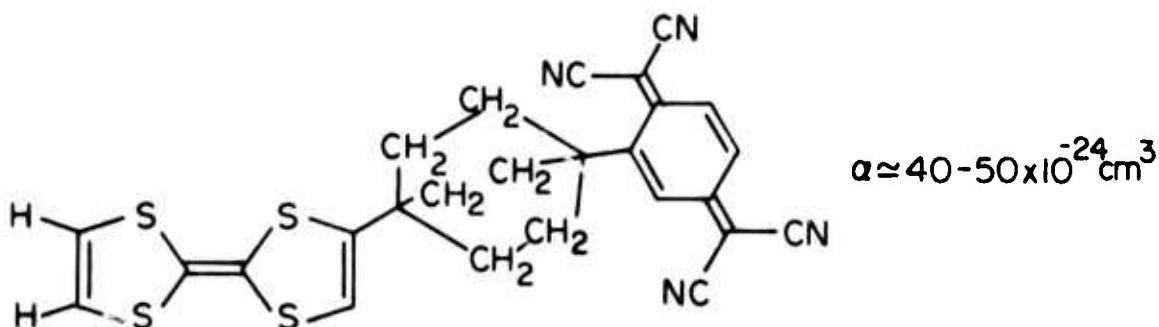
$$\text{Capacitance } C = \frac{Mdd' [1 + (4\pi\alpha' / \ell dd')]}{4\pi\ell} \quad (1)$$

$$\text{Resistance } R = \frac{M\ell\rho}{dd'} \quad (2)$$

Fraction of resistive (versus total) current flow

$$\frac{I_R}{I_{RC}} = \frac{1}{(1 + \omega^2 C^2 R^2)^{1/2}} \quad (3)$$

The hypothetical molecule $\underline{3}$ was chosen as the first molecule to be treated. The net polarizability, α' , of molecule $\underline{3}$ was estimated by vector summing the individual bond



3

polarizabilities (which were provided by W. Flygare). The length ℓ was defined as the molecule's long axis approximately 20\AA while the width d , and stacking distance d' , were approximated as 8\AA and 6\AA , respectively. Since the bulk conductivity along the X -axis (perpendicular to the stacking axis) is unknown for this molecule, it was treated as a variable. Actually, by inserting different chemical functions between the TTF and TCNQ groups, rather small changes in ℓ , d and d' would be induced while ρ_x could be varied widely. Thus, calculations to test the desirable magnitude of ρ_x as a function of frequency, ω , would provide a useful guide to the synthetic chemist.

Solving Eq. (1) to give C as a function of the number of molecules in a stack, M , gives:

$$C = 2.27 \times 10^{-9} M \text{ cm.}$$

Solving Eq. (2) for R as a function of M and ρ_x gives:

$$R = (4.8 \times 10^{-5}) (M \rho)$$

(where ρ can be substituted in as ohm cm, because the unit

correction factors to convert to esu have already been introduced). Solving Eq. (3) then gives:

$$\frac{I_R}{I_{RC}} = \frac{1}{[1+M^4 \omega^2 \rho_x^2 (1.74 \times 10^{-16})]^{\frac{1}{2}}}$$

For TTF-TCNQ $\rho_x \approx 1$ ohm cm. However, one expects the resistivity to be higher for molecule 3. Thus, allowing ρ_x to have several values (shown below) the following solutions are found.

$$\rho_x = 1 \quad \frac{I_R}{I_{RC}} = \frac{1}{[1+M^4 \omega^2 (1.74 \times 10^{-16})]^{\frac{1}{2}}}$$

$$\rho_x = 10^3 \quad \frac{I_R}{I_{RC}} = \frac{1}{[1+M^4 \omega^2 (1.74 \times 10^{-10})]^{\frac{1}{2}}}$$

$$\rho_x = 10^6 \quad \frac{I_R}{I_{RC}} = \frac{1}{[1+M^4 \omega^2 (1.74 \times 10^{-4})]^{\frac{1}{2}}}$$

$$\rho_x = 10^9 \quad \frac{I_R}{I_{RC}} = \frac{1}{[1+M^4 \omega^2 (1.74 \times 10^2)]^{\frac{1}{2}}}$$

Under these conditions, the value of I_R/I_{RC} is reasonable for a variety of frequencies, as long as M (the number of molecules in an array) is small. Consider Table I where M is assumed to be small. As long as M is kept small, it appears that a useful rectifier is possible for moderately low frequencies. It is important to have low values of ρ_x if higher frequencies are considered. Thus, the chemist is faced with the task of designing molecules with fairly high values of ρ_x , which can then be aligned between end plates in small groups. This concept appears valuable for low frequency applications.

TABLE I. I_R/I_{RC} as a function of M , ω and ρ_x for a Molecular Rectifier.

ω_{cgs}	M	ρ_x ohm cm.	I_R/I_{RC}
10^4	10	1	$1/[1+1.74 \times 10^{-1}]^{\frac{1}{2}}$
10^7	10	1	$1/[1+1.74 \times 10^2]^{\frac{1}{2}}$
10^{10}	10	1	$1/[1+1.74 \times 10^8]^{\frac{1}{2}}$
10^4	10	10^3	$1/[1+1.74 \times 10^2]^{\frac{1}{2}}$
10^7	10	10^3	$1/[1+1.74 \times 10^8]^{\frac{1}{2}}$
10^{10}	10	10^3	$1/[1+1.74 \times 10^{14}]^{\frac{1}{2}}$
10^4	10	10^6	$1/[1+1.74 \times 10^8]^{\frac{1}{2}}$
10^7	10	10^6	$1/[1+1.74 \times 10^{14}]^{\frac{1}{2}}$
10^{10}	10	10^6	$1/[1+1.74 \times 10^{20}]^{\frac{1}{2}}$
10^4	1	1	$1/[1+1.74 \times 10^{-8}]^{\frac{1}{2}}$
10^{10}	1	1	$1/[1+1.74 \times 10^4]^{\frac{1}{2}}$
10^4	1	10^3	$1/[1+1.74 \times 10^{-2}]^{\frac{1}{2}}$
10^7	1	10^3	$1/[1+1.74 \times 10^4]^{\frac{1}{2}}$
10^4	1	10^6	$1/[1+1.74 \times 10^4]^{\frac{1}{2}}$
10^7	1	10^6	$1/[1+1.74 \times 10^{10}]^{\frac{1}{2}}$
10^4	10^2	10^3	$1/[1+1.74 \times 10^6]^{\frac{1}{2}}$

ACKNOWLEDGEMENT

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-71-C-033 with The University of Michigan.